

XX 12-SEP-2001; 2001WO-US28488.
 XX 11-SEP-2000; 2000US-0659671.
 PR 06-SEP-2001; 2001US-317902P.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT;
 PI WPI; 2002-393966/42.
 DR Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects
 XX Disclosure; Page 131-132; 152pp; English.
 PS
 CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy2
 CC mature protein.
 XX
 SQ Sequence 365 AA;
 Alignment Scores:
 Pred. No.: 4,72e-152 Length: 365
 Score: 1750.50 Matches: 352
 Percent Similarity: 96.73% Conservative: 3
 Best Local Similarity: 95.91% Mismatches: 8
 Query Match: 44.10% Indels: 7
 DB: 23 Gaps: 1
 US-10-060-830-2 (1-2190) x AAE22721 (1-365)
 QY 61 CAGCAAGTGATGGTGGACACACTGTACTAGGCCCTGAGAGTGGAAACCTTACATCC 120
 Db 1 GlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGlnSerGlyThrLeuThrSer 20
 QY 121 ATAACTACCCAGACACTTCCCAACAGCACACTGTTTGTGAATGGGAGATCCCGTAAAG 180
 Db 21 IleAsnTyrProGlnThrTyrProAsnSerThrValCysGlnTrpGluIleArgValLys 40
 QY 181 ATGGGAGAGAGTTCGCATCAATTTGGTGCACCTTGAATTTGATCTGATCTTGT 240
 Db 41 MetGlyGluArgValArgIleLysPheGlyAspPheAspGluAspSerAspSerCys 60
 QY 241 CACTTTTACTTACGAAATTTATATGGAATTCGAGTTCAGCAGCACTGAAATAGGAAA 300
 Db 61 HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys 80

QY 301 TACTCTGGTCTGGGGTTCGAAATGAACCAATTCATTAATGAATCAAAAGCGCAATGAATCACA 360
 Db 81 TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr 100
 QY 361 TTGCTGTTTCATGAGTGGAAATCCATGTTTCTGGACGGGGATTTTGGCCCTCATACTCTGT 420
 Db 101 LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal 120
 QY 421 ATAGATAAACAAAGATCTAATTTACTTTTGGACACTGCATCCCAATTTTGGAACTCTGAG 480
 Db 121 IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu 140
 QY 481 TTCAGTAAGTACTGCCACCTGGTGTCTGCTCTCTCTTCTGCTGAGATATCTGGAACAATT 540
 Db 141 PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIle 160
 QY 541 CCTCATGGATATAGATTCCTCGCCATTTGTCATGGCTGGTGTGCATGCAGAGTAGTG 600
 Db 161 ProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValVal 180
 QY 601 TCAACACAGCTTGGGGCGGCCAAATCACTGTGTAAATAGTAAAGGTATTCCTTATTATGAA 660
 Db 181 SerAsnThrLeuGlyGlyGlnIleSerValIleSerLysGlyIleProTyrTyrGlu 200
 QY 661 AGTCTCTTGGCTAACAACTGCATCTGTGGTGGGACACTTATCTACAAGTCTTTTATACA 720
 Db 201 SerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr 220
 QY 721 TTTAAGACAAAGTGGATGTTTATGGAACACTGGGATGGAGTCTGGTGT-GATCGCGGATCC 779
 Db 221 PheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyGlyAspArgGlySer 240
 QY 780 TCAATAACAGCATCATCTGCTGGTGGAGTGGATGACACACAGGGCAAGAGACAGTTG 839
 Db 241 SerAsnAsnSerIle-ThrValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTr 260
 QY 840 GAACCCCAAAAGC-CAGGCTGAAAAAACCCTGGACCGCCTGGGCTGCTTTTGGCCACTG 898
 Db 260 PysProLysLysSerGlnAlaGluLysThrTrpThrAlaLeuGly-AlaPheAlaThrA 280
 QY 899 ATGAATACCAAGTGGTTTACAAATAGATTGAATAGGAAAAAGAAATAACAGGCATTATAA 958
 Db 280 spGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIle 300
 QY 959 CCACCTGGATCCACCATGGTGGAGCACAAATFACATGTGTGTCTGCTCAGCAATCTCTGTACA 1018
 Db 300 hrThrGlySerThrMetVal-SerThrIleThrMetCysLeuProThrGluSerCysThr 319
 QY 1019 GTGATGATGGGCAAGAAATGGACTGTGTACAGAGAGCCCTGGT-GTGGAGCAAGATAAGATA 1077
 Db 320 ValMetMetGlyArgAsnGlyLeuCysThrGluSerLeuValValGluGlnAspLysIle 339
 QY 1078 TTTCAAGGAAACAAAGAT---TATCACACGAGGTGTGGGTAACTAACTTTTGGCCACCAATT 1134
 Db 340 PheGlnGlyAsnLysArgIleIleThrArgMetValArgAsnAsnPheLeuProPheIle 359
 QY 1135 ATTGCACGCTTTTATT 1149.
 Db 360 IleAlaArgLeuLeu 364
 RESULT 7
 AAU00670
 ID AAU00670 standard; Protein; 715 AA.
 XX
 AC AAU00670;
 XX
 XX 07-SEP-2001 (first entry)
 DT
 XX Human TANGO 229 polypeptide.
 DE
 XX Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
 KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;

Db 370 GlnGlyAsnSerAspPheArgAspProValGlnAsnAsnPhleProProlleValala 389
QY 1141 CQTATTATTAGTGAATCCTACCAATGGCAGCAGAAATTCATGCAATGAGAGTG 1200
Db 390 ArgTyrValArgValProGlnThrPheGlnArgGlnAlaLeuLysValGluLeu 409
QY 1201 CTCGATGTCAGTTATTCCTAAAGGTGCTCTCCAAAACCTTACTCAACCTCCACTCCT 1260
Db 410 IleGlyCysGln-----IleThrGln----- 416
QY 1261 CGGAACAGCAATGAC-----CTCAAAAACACTACAGCCCTCCAAAATAGCC 1308
Db 417 -----GlyAsnAspSerLeuValTrpArgLysThrSerGlnSerThrSerValSerThr 434
QY 1309 AAAGTGTGTCGCCAAAATTTACGAAACACACTACAACTCGCAGTAGCAATGAATTCCT 1368
Db 435 LysLysGluAspGluThrIleThrArgProle----- 445
QY 1369 GCACAGACAGCAAACTCCAGTCTGATATCAGAACTACTACGTAACCTCAAT 1428
Db 446 -----ProSerGluGluThr-----SerThrGlyIleAsnIleThr 458
QY 1429 GTAACCAAGATGAGCGTGTGCTGCAGTTCTTGTCCCTGTGCTGCTCATGCTCCTCACT 1488
Db 459 -----ValAlaIleProLeuValLeuValValPheAlaGlyMet 474
QY 1489 ACTCTCATCTCATATTAGTGTGTGCTGGCAGTGGCACTGAGAAACAGAAAGAAAACCTGAA 1548
Db 475 GlyIlePheAlaAlaPhe-----ArgLysLysLysLys 486
QY 1549 GCACCTATGACTTACCTTACTTGG-----GACCGGCAGGTGGTGAAGGA 1596
Db 487 GlySer-----ProTyrGlySerAlaGluAlaGlnLysThrAspCysTrpLysGln 503
QY 1597 ATGACAGCAGTTCTCTCTGCAAAAGCAGTGGACCATGAGAAACCCCA---GTTGCGCAT 1653
Db 504 IleLysTyr-----ProPheAlaArgHis 511
QY 1654 ACAGCAGCGAA-----GTTAATCAGTGTGATCAGAAAGACTCACC-----ACA 1698
Db 512 GlnSerAlaGlnPheThrIleSerTyrAspAsnGluLysGluMetThrGlnLysLeuAsp 531
QY 1699 GTGCTCAGCGTACCTCTCAGAGATGCTCAGCCACTGTAGGAGGATTTGGTACA 1758
Db 532 LeuIleThrSerAspMetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrVal 551
QY 1759 CTTCAATCAAGATCTACCTTTAAACCA-----GAAGAAGGAAAGAACAGCGCTATGCA 1812
Db 552 ThrArgLysGlySerThrPheArgProMetAspThrAspAlaGluGluAlaGly---Val 570
QY 1813 GACCTAGATCCT-----TACAACTCAGCGGCGAGGAGTTATCATCCCTATGCT 1863
Db 571 SerThrAspAlaGlyGlyHisTyrAspCysProGlnArgAlaGlyArgHisGluTyrAla 590
QY 1864 GAACCACTCCCAATTCAGGGCTGAGTATGCAACCCCAATCATCATGACATGTGTCAGG 1923
Db 591 LeuProLeuAlaProProGluProGluTyrAlaThrProIleVal-----GluArg 607
QY 1924 CACCCCAACACTTCACTTGGTTCAGCCCTCCACCACTCACTTTCAAGGCTACGGGGAACCAA 1983
Db 608 HisValLeuArgAlaHisThrPheSerAlaGlnSerGlyTyrArgValProGlyProGln 627
QY 1984 CTT-----CCCCCACTA-----GTGGGAAC 2004
Db 628 ProGlyHisLysHisSerLeuSerSerGlyPheSerProValAlaGlyValGlyAla 647
QY 2005 TACAATACACTTCTCCAGGACTGACAGTCTCTCTCAGCCAGCCAGCCAGTATGATACC 2064
Db 648 GlnAspGlyAspTyrGlnArgProHisSerAlaGlnProAlaAspArgGlyTyrAspArg 667
QY 2065 CCGAAAGCT 2073
Db 668 ProLysAla 670

RESULT 8

ABG01298
ID ABG01298 standard; Protein: 197 AA.

AC ABG01298;

XX 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #1289.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR N-PSDB; AAS65485.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 20; SEQ ID NO 31657; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: the sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 197 AA;

Alignment Scores:

Pred. No.: 9,8e-87 Length: 197
Score: 1039.00 Matches: 197
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.18% Indels: 0
DB: Gaps: 0

US-10-060-830-2 (1-2190) x ABG01298 (1-197)


```
QY 613 GCGGCGCAATCAGTGTGTTAATAGTAAGATATCCCTATATATGAAGTCTTTGGCT 672
Db 214 GlyGlnIleSerValLeuGlnArgLysGlyIleSerArgThrGluGlyIleLeuAla 233
QY 673 AACAGGTCACATCTGTGTGGACATATATCTACAAGTCTTTTACATTTAAAGACAAGT 732
Db 234 AsnGlyValLeuSerArgAspGlySerLeuSerAspLysArgPheLeuPheThrSerAsn 253
QY 733 GGATGTTATGAACATCGGGGATGGAGTCTGTGTGATCGCGGATCCTCAAAATACAGCA 792
Db 254 GlyCysSerArgSerLeuSerPheGlu-----ProAspGlyGlnIleArgAla 269
QY 793 TCATCTGTCTGGATGGAGTGCACACACAGGCGAAGACAGATGGAACCCAAAAA 852
Db 270 SerSerTrpGlnSerValAsnGluSerGlyAspGlnValHisTrpSerProGlyGln 289
QY 853 GCCAGGCTGAAAAACCTGGACCGCTTGGCTGCTTTTCCCACTGATGAA----- 903
Db 290 AlaArgLeuGlnAspGlnGlyProSerTrpAlaSerGlyAspSerSerAsnAsnHisLys 309
QY 904 ---TACCAGTGTGTACAAATAGATTGAATAAGAAAGAAATAACAGGCATTATAACC 960
Db 310 ProArgGluTrpLeuGluIleAspLeuGlyGluLysLysIleThrGlyIleArgThr 329
QY 961 ACTGGATCCACCATGTGGACACAAATCTACTGTGTCTGCCTACAGAACTCTGTACAGT 1020
Db 330 ThrGlySerThrGlnSerAsnPheAsnPheTyrrValLysSerPheValMetAsnPheLys 349
QY 1021 GATGATGGGAGAAATGGATGTGTACAGAGAGCCGTGTGGAGCAAGATAGATATT 1080
Db 350 AsnAsnAsnSerLysTrpLysThrTyrrLysGlyIleValAsnGluGluLysValPhe 369
QY 1081 CAAGGAACAAGAAATATCACCAGGATGTGCGGTAACTATTTTGGCCCACTAATTTGCA 1140
Db 370 GlnGlyAsnSerAsnPheArgAspProValGlnAsnAsnPheIleProIleValAla 389
QY 1141 CGTTTATAGATGATCTACCAATGGCAGCAAAATGCCATGAAATGGAGCTG 1200
Db 390 ArgTyrrValArgValProGlnThrTrpHisGlnArgIleAlaLeuLysValGluLeu 409
QY 1201 CTGGATGTGAGTTTATCTAAAGTCGTCCTCAAAACTTACTCAACTCCACCTCCT 1260
Db 410 IleGlyCysGln-----IleThrGln----- 416
QY 1261 CGGAACAGCAATGAC-----CTCAAAACACTACAGCCCTCCAAATAAGCC 1308
Db 417 -----GlyAsnAspSerLeuValTrpArgLysThrSerGlnSerThrSerValSerThr 434
QY 1309 AAGGTGCTGCCCAAAATTTAGCAACCACTACACCTCGCAGTAGCAATGATTCCT 1368
Db 435 LysLysGluAspGluThrIleThrArgProIle----- 445
QY 1369 GCACAGACAAACAACAACCTGCAGTCCTGATATCAGAAATACCTACCTCAAAAT 1428
Db 446 ---ProSerGluGluThr-----SerThrGlyIleAsnIleThr----- 458
QY 1429 GTAACCAAGATGAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1488
Db 459 -----ValAlaIleProLeuValLeuValLeuValValPheAlaGlyMet 474
QY 1489 ACTCTATCTCTATATAGTGTGTGCTGGCACTGGAGAAACAGAAAGAAACTGAA 1548
Db 475 GlyIlePheAlaAlaPhe-----ArgLysLysLysLysLys 486
QY 1549 GGCACCTATGACTTACTTACTGG-----GACCGGCGAGTTGGTGAAGGA 1596
Db 487 GlySer-----ProTyrrGlySerAlaGluAlaGlnLysThrAspCysTrpLysGln 503
QY 1597 ATGAAGCAGTTTCTCTCTGCAAAAGCAGTGGACCATGAGGAAACCCCA-----GTTCCGTAT 1653
Db 504 IleLysThr-----ProPheAlaArgHis 511
```

1654 AGCAGCGGAA-----GTTAATCACCTGAGTCCCAAGAGAAGTCCACACAGTGTGCAG 1707
Db 512 GlnSerAlaGluPheThrIleSerTyrrAspAsnGluLysGluMetThrGlnLysLeuAsp 531
QY 1708 GCTGACTCTGCAGAGATGCT 1728
Db 532 LeuIleThrSerAspMetAla 538

RESULT 10
AAU00629
ID AAU00629 standard; Protein; 586 AA.
XX
AC AAU00629;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #2.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW Obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
DR WPI: 2001-290917/30.
DR N-PSDB; AAS00614.
XX
PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
PS Claim 2; Page 27-28; 33pp; English.

The sequence represents a novel human protein (NHP) containing a CUB domain (an extracellular domain). CUB proteins have been associated with regulating development, modulating cellular processes and preventing infectious disease. NHP nucleotide sequences are useful for gene therapy of physiological disorders or diseases. NHP oligonucleotides are useful as hybridisation probes for screening libraries and assessing gene patterns. NHP nucleotide sequences are useful for detecting mutant or inappropriately expressed NHPs (for example, those proteins associated with obesity, high blood pressure, connective tissue disorders and infertility) for the diagnosis of a disease. The polynucleotides may also be used in screening for drugs effective in the treatment of symptomatic or phenotypic manifestations of perturbing the normal function of NHP in the body. Nucleotide constructs encoding NHP products are used to genetically engineer host cells to express such products in vivo. These host cells allow for the identification of compounds that bind to NHP receptors or trigger NHP-mediated pathways.

Alignment Scores:
Pred. No.: 5,15e-76 Length: 586
Score: 924.50 Matches: 221
Percent Similarity: 54.34% Conservative: 98
Best Local Similarity: 37.65% Mismatches: 185
Query Match: 23.29% Indels: 83
Gaps: 17
SQ Sequence 586 AA;

(LEXI-) LEXICON GENETICS INC.
Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
WPI; 2001-290917/30.
N-PSDB; AAS00613.

Novel nucleic acid encoding human CUB-domain containing protein, useful
for drug screening, diagnosis and treatment of physiological disorders
or diseases

Claim 4; Page 26; 33pp; English.

The sequence represents a novel human protein (NHP) containing a CUB
domain (an extracellular domain). CUB proteins have been associated with
regulating development, modulating cellular processes and preventing
infectious disease. NHP nucleotide sequences are useful for gene therapy
of physiological disorders or diseases. NHP oligonucleotides are useful
as hybridization probes for screening libraries and assessing gene
patterns. NHP nucleotide sequences are useful for detecting mutant or
inappropriately expressed NHPs (for example, those proteins associated
with obesity, high blood pressure, connective tissue disorders and
infertility) for the diagnosis of a disease. The polynucleotides may also
be used in screening for drugs effective in the treatment of symptomatic
or phenotypic manifestations of perturbing the normal function of NHP in
the body. Nucleotide constructs encoding NHP products are used to
genetically engineer host cells to express such products in vivo. These
host cells allow for the identification of compounds that bind to NHP
receptors or trigger NHP-mediated pathways.

Alignment Scores:		
Pred. No.:	3.67e-70	Length:
Score:	860.50	Matches:
Percent Similarity:	54.33%	Conservative:
Best Local Similarity:	37.18%	Mismatches:
Query Match:	21.68%	Indels:
DB:	22	Gaps:
		17
US-10-060-830-2 (1-2190) x AAU000628 (1-487)		487

Db 243 ----- 243
 QY 1024 GATGGCAGAAATGGACTGTGTACAGAGAGCGTGGTGGGACAGATAAGATATTCTCA 1083
 Db 244 ----- 246
 QY 1084 GGAACAAAGATTATCACCGAGGTGGCGTAATACTTTTCCACCAATTATTGCACGT 1143
 Db 246 ----- 246
 QY 1144 TTTATTAGAGTGAATCTTACCAATGGCAGAGAAATTTGCCATGAAATGGAGTGTCTC 1203
 Db 246 ----- 246
 QY 1204 GGATGTCAGTTTATCTTAAAGTCTCTCTCCAAACTTACTTCARCTCCACCTCTCTCGG 1263
 Db 246 ----- 246
 QY 1264 AACAGCAATGACCTCAAAACACTACAGCGCCCTCCAAATAAGCCAAAGGTGGTCCCA 1323
 Db 246 ----- 246
 QY 1324 AAATTTAGCAACCACTACAACTCGCAGTAGCAATGAATTTCTCGCAGACAGAACAA 1383
 Db 246 ----- 246
 QY 1384 ACACTGCCAGTCTGTATATCAGAAATACTACGTAACCTCCAAATTAACCAAGATGTA 1443
 Db 247 ----- 251
 QY 1444 CGCGTGGCTGCACTTTCTGTCCCTGTGGTGCATGGTCTCTACTACTCTCTCTCTCA 1503
 Db 252 AlaileProSerValile-----PheileAlaileLeuLeuThrGlyMetGlyIlePhe 269
 QY 1504 TTAGTGTGTCTGGCACTGGAGAAACAGAAAGAAAACTAAGCGCCCTATGACTTA 1563
 Db 270 AlaileCys-----ArgLysArgLysLysLysGlyAsn----- 280
 QY 1564 CTTTAC-----TGGGACCGGCGAGTGTGGTGGAAAGGAATGAAGCAGTTTCTT 1611
 Db 281 ProTyrValSerAlaAspAlaGlnLysThrGlyCysTrpLysGlnIleLysTyr----- 298
 QY 1612 CCTGCAAAACGCTGGACCATGAGGAACCCCA---GTTGCTATAGCAGCAGCAA--- 1665
 Db 299 ----- 308
 QY 1666 ---GTTAATCACTGAGTCCACAGAGAAGTCACC-----ACAGTGTCTGCAGGCTGAC 1713
 Db 309 ThrIleSerTyrAspAsnGlnLysGluMetThrGlnLysLeuAspLeuIleThrSerAsp 328
 QY 1714 TCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTGTACACTTCATCAAGATCT 1773
 Db 329 MetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrValAlaArgLysGlySer 348
 QY 1774 ACCTTTAAACA-----GAGAA-----GGAAAGAGCAGCG 1806
 Db 349 ThrPheArgProMetAspThrAspThrGluGluValArgValAsnThrGluAlaSerGly 368
 QY 1807 TATGCAGACCTAGATCTTCACTACACTCACCAGGCGGAGAGTTTATCATGCTATGTGAA 1866
 Db 369 His-----TyrAspCysProHisArgProGlyArgHisGluTyrAlaLeu 383
 QY 1867 CCACTCCCAATTAGGGGCTGAGTATGCAACCCCAATCATC----- 1908
 Db 384 ProLeuThrHisSerGluProGluTyrAlaThrProIleValGluArgHisLeuLeuArg 403
 QY 1909 -----ATGGACATGTCAGGCCACCCCAACTTCAGTTGGTGCAGCCCTCCACA 1956
 Db 404 AlaHisThrPheSerThrGlnSerGlyTyrArgValProGly---ProArgProThrHis 422
 QY 1957 TCCACTTTCAGGCTACGGGACCAACCTCCCTCCACTTCTGGGAATTCACATACACTT 2016

Db 423 GluHisSerHisSerSerGlyGlyPheProAlaThrGlyAlaThrGlnValGluSer 442
 QY 2017 CTCTCCAGGACTGACAGCTGCTCCTCAGCCAGCCAGTATGATACCCCGAAAGCTGGG 2076
 Db 443 TyrGlnArgProAlaSerProLysProValGlyGlyTyrAspLysPro---AlaAla 461
 QY 2077 AAGCCAGTCTTACCTGCCCGCAGACGAATTTGGTGTACCGAGTGCACAGACACAAAGAA 2136
 Db 462 SerSerPheLeuAspSerArgAsp-----ProAlaSerGlnSerGln 475
 QY 2137 GTATCAGGACGAGGAGGATGGG 2160
 Db 476 MetThrSerGlyGlyAspAspGly 483
 RESULT 13
 AAE22715
 ID AAE22715 standard; Protein; 398 AA.
 XX
 AC AAE22715;
 DT
 DT 09-AUG-2002 (first entry)
 XX
 DE Human neuropilin-Hy1 protein.
 XX
 KW Human; neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
 KW immunosuppressive; chromosome 6q21.
 XX
 OS Homo sapiens.
 XX
 PN WO200222815-A1.
 XX
 PD 21-MAR-2002.
 XX
 PF 12-SEP-2001; 2001WO-US28488.
 XX
 PR 11-SEP-2000; 2000US-0659671.
 PR 06-SEP-2001; 2001US-317902P.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT;
 XX
 DR WPI: 2002-393966/42.
 DR N-PSDB; AAD35992.
 XX
 PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects
 XX
 PS Claim 3; Page 123-125; 152pp; English.
 XX
 CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat

CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-brager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hyl
 CC protein. Neuropilin-Hyl gene is located on chromosome 6q21.

XX
 SQ Sequence 398 AA;

Alignment Scores:
 Pred. No.: 9e-41 Length: 398
 Score: 540.50 Matches: 138
 Percent Similarity: 45.84% Conservative: 55
 Best Local Similarity: 32.78% Mismatches: 117
 Query Match: 13.62% Indels: 111
 DB: 23 Gaps: 11

US-10-060-830-2 (1-2190) x AA22715 (1-398)

QY 64 CAAGGTGATGGATGTGCACACACTGACTAGGCGCTGAGAGTGGAGCCCTTACATCCATA 123
 DB 3 GUGLYASPGLYCYSGLYHISLEUVALTHRTYRGLNASPSERGLYTHRMETHSERLYS 22
 QY 124 AACTACCCACAGACCTATCCACAGACACTGTTTGTGAATGGAGATCCGTGTAAAGATG 183
 DB 23 ASNTYRPROGLYTHRTYRPROASNHSITRVALCYSGLYSLYTHRIETHRVALPROLYS 42
 QY 184 GGAGAGAGTTCGCATCAATTTGGTACTTTGACATTTGACATTTGATCTTGTTCAC 243
 DB 43 GYLYARGLEULEULEUARGLEUGLYASPLEUASPLEUGLU---SERGLNTHRCYSALA 61
 QY 244 TTTAATTAATCTAGAGATTTATATGGAATGGAGTACAGACACTGAATAGGCAATATAC 303
 DB 62 SERASPTYRLEULEUPHEHRSER-----SERSERASPGINTYRGLYPROTYR 77
 QY 304 TGTGCTGGGGTGGCAATGAACCATTAATGAATCAAAAGCAATGAATACACATTTG 363
 DB 78 CYSGLY---SERMETHTRVALPROLYSGLULEULEUASNTHSERGLUVALTHRVAL 96
 QY 364 CTGTTATGAGTGAATCCATGTTCTGGACGGGATTTTGGCTCATCTCTGTTATA 423
 DB 97 ARGPHLEULSERGLYSERHISLEISERGLYARGGLYPHELEULEUHTYRVALASER 116
 QY 424 GAT-----AAACAAGATCTAATT 441
 DB 117 ASPHISPROGLUSERGLINGLYASPTGPROSERGLULYSTRHLEUASPGINGLINSER 136
 QY 442 ACTTGTGTTGACACTGCATCCAAATTTTGGAACTGAGTTACAGTACAGTACGCCAGCT 501
 DB 137 THRPHLEULALATHRGLYTHRTYRPHLEVALLYASPSERPHESER----- 151
 QY 502 GGTGCTGCTCTCTTTGCTGAGATATCTGGAACATTTCTCATGATATAGAGATCC 561
 DB 152 -----ThrAspGlyThr----- 155
 QY 562 TCGCAFTTGTGCTGGTGTGTCAGAGTACTGCAACAGCTTGGCGGCCAA 621
 DB 156 SERLEULEUCYSGLYALAALAIHISALAGLYLEILEALASPGLEUGLYGLN 175
 QY 622 ATCAGTCTTGAATTAAGGATTTCCCTATATGAAGATTTCTTGGCTTAACAACGTC 681
 DB 176 ILESERVALLEUGLNASRGLYSGLYILESERARGTYRGLUGLYILELEUALAASNGLYVAL 195

QY 682 ACATCT----- 687
 DB 196 LEUSERARGGLUPHEGLUILEPHEARGGLULNLEUPHESERVALLEUPHEFYR 215
 QY 688 -----GTGGTGGGACAC 699
 DB 216 TRPGLYASNTHRVALHISALAVALLIEGLULEUMETPHEPROHISMETILEVALTRPHIS 235
 QY 699 ----- 699
 DB 236 SERGLYLYSTRHARGGLUGLYSERILEALALAGLUGLUGLYVALPROLYSLEUTYR 255
 QY 700 -----TTATCTACAAGTCTTTTACATTTTAAGACAGTGGATGT 738
 DB 256 LEUVALILEGLNLYSGINGLULEUVALGLNASPLEUVALLEUVALALATHRVALGLYCYS 275
 QY 739 TATGGAAACACTGGGATGGAGTCTGTGTGATCGCGGATCTCAATATAACAGCATCATCT 798
 DB 276 SERARGSERLEUSERPHEGLU-----PROASPGLYGLNILEARGALASER 291
 QY 799 GTCTGGAGTGGACTGACACACAGAGCAAGAGCAAGTGTGGAAACCCAAAAAGCCAGG 858
 DB 292 SERTRPGINSERVALASNGLUSERGLYASPGINVAIHISTRPSERPROGLYGLNIALARG 311
 QY 859 CTGAAATAAACCCTGGACCGCTTGGCTGTCTTTGCCACTGATGAA-----TAC 906
 DB 312 LEUGLNASPGINGLYPROSERTRPAALASERGLYASPSERSERASNHSILYSPROARG 331
 QY 907 CAGTGGTTACAAATAGATTTGAATAAGAAAGAAATAACAGCAATTAACACATGGA 966
 DB 332 GLUTRPLEUGLUILEASPLEUGLYGLULYSLSYLSILETHRGLYILEARGTHRGLY 351
 QY 967 TCACACATGTTGGAGCACAATTTACTATGTTCTGCTCCACAGATCTCTGACATGAT 1026
 DB 352 SERTHRGLINSERASNPHESASNPHETRYVALYSSERPHETVALMETASNPHELYSASN 371
 QY 1027 GGCAGAAATGGAGTGTGTACAGAGAGCCCTGGTGTGCGCAAGATAAGATATTCAAGGA 1086
 DB 372 ASNSERLYSTRIPLYSTRYRGLYSGLYILEVALASNANGLUGLULYSVAL---ARGGLY 390
 QY 1087 AAC 1089
 DB 391 ASN 391
 RESULT 14
 AAU79459
 ID AAU79459 standard; Protein; 398 AA.
 XX
 AC AAU79459;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Human Neuropilin-Hyl.
 XX
 KW Human; neuropilin-Hyl; chromosome 6q21; neuronal growth;
 KW nerve regeneration; neurodegenerative disease; learning disorder;
 KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
 KW organ growth; nervous system lesion; cancer; cell proliferation;
 KW cell differentiation; stem cell growth factor activity;
 KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
 KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
 KW platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
 KW reperfusion; food supplement; DNA microarray.
 XX
 OS Homo sapiens.
 XX
 PN WO200222780-A2.
 XX
 PD 21-MAR-2002.
 XX
 PF 11-SEP-2001; 2001WO-US28590.
 XX

PR 11-SEP-2000; 2000US-0659671.
 PR 06-SEP-2001; 2001US-0659671.

(TANG/) TANG T Y.

XX Tang TY;

DR WPI; 2002-351881/38.

DR N-PSDB; ABK49565.

XX New neuropilin-like polypeptides for diagnosing, preventing and
 PT treating neurological conditions and disorders, cancers, and for
 PT inducing angiogenesis and neovascularisation

XX Claim 3; Page 118-120; 144pp; English.

XX The invention relates to an isolated neuropilin-like polypeptide
 CC including neuropilin-Hyl and neuropilin-Hy2, the full length cDNAs
 CC encoding the proteins and the coding regions of the cDNAs. Also included
 CC is a nucleic acid array comprising the cDNAs attached to a surface used
 CC for detecting full-matches or mismatches to the cDNAs. The genes
 CC for neuropilin-Hyl and Hy2 are located on human chromosome 6q21.
 CC The nucleic acid array is useful for detecting full-matches or mismatches
 CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
 CC in modulating neuronal growth, regenerative capacity, treating
 CC neurodegenerative diseases, learning and memory disorders, diagnosing and
 CC mapping genetic neuronal defects and degenerative diseases like
 CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
 CC and organ growth and development (e.g. the heart). The nervous system
 CC disorders include lesions of central or peripheral nervous systems,
 CC including traumatic lesions, ischaemic lesions, infectious diseases,
 CC degenerative lesions, lesions associated with nutritional diseases or
 CC disorders, neurological lesions, and lesions caused by toxic substances.
 CC The neuropilin-like proteins and cDNAs are also useful as markers for
 CC cancers. The neuropilin-like proteins are useful for regulating cell
 CC proliferation, cell differentiation, stem cell growth factor activity,
 CC for inducing proliferation of neural cells, regeneration of nerve and
 CC brain tissue, for treatment of central and peripheral nervous system
 CC diseases, and neuropathies, such as Parkinson's disease,
 CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
 CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
 CC anaemias, and platelet disorders, such as thrombocytopaenia,
 CC regeneration and treatment of lung or liver fibrosis, reperfusion
 CC injury in various tissues and as a food supplement or molecular
 CC weight marker. The cDNAs are useful in gene identification, genome
 CC mapping, transgenics, as hybridisation probes, for primer design, for
 CC gene chips and as a DNA antigen. The present sequence represents
 CC neuropilin-Hyl.

XX SQ Sequence 398 AA;

Alignment Scores:
 Pred. No.: 9e-41 Length: 398
 Score: 540.50 Matches: 138
 Percent Similarity: 45.84% Conservative: 55
 Best Local Similarity: 32.78% Mismatches: 117
 Query Match: 13.62% Indels: 111
 DB: 23 Gaps: 11

US-10-060-830-2 (1-2190) x AAU79459 (1-398)

QY 64 CAAGGTGATGGATGGACACACTGTACTAGGCCCTGAGAGTGGACCCCTACATCCATA 123
 :::::::::::::::::::::: ::::::::::::::::::::::
 Db 3 GluGlyAspGlyCysGlyHisLeuValThrTyrglnAspSerGlyThrMetThrSerLys 22
 :::::::::::::::::::::: ::::::::::::::::::::::
 QY 124 AACTACCCACAGACTTCCCAACAGACACTGTTTGTGAATGGGAGATCCGTGTAAGATG 183
 :::::::::::::: ::::::::::::::
 Db 23 AsnTyrProGlyThrTyrProAsnHisThrValCysGlnLysThrIleThrValProLys 42
 :::::::::::::: ::::::::::::::
 QY 184 GGAGAGAGAGTCCCATCAATTTGGTGACTTTCACATTTAGAGATTCTGATCTTGTTCAC 243
 :::::::::::::: ::::::::::::::
 Db 43 GlyLysArgLeuIleLeuArgLeuGlyAspLeuAspIleGlu---SerGlnThrCysAla 61
 :::::::::::::: ::::::::::::::

QY 244 TTTAATTACTTGAGATTATTAATGAATTGGAGTCAGCAGAACTGAATAGCAATATC 303
 :::::::::::::: ::::::::::::::
 Db 62 SerAspTyrLeuLeuPheThrSer-----SerSerAspGlnTyrGlyProTyr 77
 :::::::::::::: ::::::::::::::
 QY 304 TGTGTCTCTGGGTTGCCAAATGAACCACTCAATTCAATCAAAAGGCAATGAATACACATTG 363
 :::::::::::::: ::::::::::::::
 Db 78 CysGly---SerMetThrValProLysGluLeuLeuAsnThrSerGluValThrVal 96
 :::::::::::::: ::::::::::::::
 QY 364 CTGTTCATGAGTGAATCCATGTTCTGGACGCGGATTTTGGCCCTCATCTCTGTATATA 423
 :::::::::::::: ::::::::::::::
 Db 97 ArgPheGluSerGlySerHisIleSerGlyArgGlyPheLeuLeuThrTyrAlaSerSer 116
 :::::::::::::: ::::::::::::::
 QY 424 GAT-----AAACAAGATCTAATT 441
 ::::::::::::::
 Db 117 AspHisProGluSerGlnGlyAspArgProSerGluLysThrLeuAspGlnSerArg 136
 ::::::::::::::
 QY 442 ACTTGTGTTGGACACTGCATCCCAATTTTGGAACTGAGTTCAGTAAGTACTGCCAGCT 501
 :::::::::::::: ::::::::::::::
 Db 137 ThrPheLeuAlaThrGlyThrThrPheValLysAspSerPheSer----- 151
 ::::::::::::::
 QY 502 GGTGTCTGCTTCTTCTTGTGATATCTGGAACTTCCTCATGATATATAGATTC 561
 ::::::::::::::
 Db 152 -----ThrAspGlyThr----- 155
 ::::::::::::::
 QY 562 TCGCCATTGTCATGCTGCTGTCATGCAGGAGTAGTCTCAACACGTTGGCGGCCAA 621
 ::::::::::::::
 Db 156 SerLeuLeuCysLysAlaIleHisAlaGlyIleIleAlaAspGluLeuGlyGlyGln 175
 ::::::::::::::
 QY 622 ATCAGTGTGTAATTAGTAAAGGTATTCCCTATTATCAAAAGTTCTTTGGGTAAACAAGTC 681
 ::::::::::::::
 Db 176 IleSerValLeuGlnArgLysGlyIleSerArgTyrGluGlyIleLeuAlaAsnGlyVal 195
 ::::::::::::::
 QY 682 ACATCT----- 687
 ::::::::::::::
 Db 196 LeuSerArgGluPheGluIlePheArgGluGlnLeuPheSerSerValLeuPheTyrSer 215
 ::::::::::::::
 QY 688 -----GTGGTGGGACAC 699
 ::::::::::::::
 Db 216 TrpGlyAsnThrValHisAlaValIleGluLeuMetPheProHisMetIleValTrpHis 235
 ::::::::::::::
 QY 699 ----- 699
 ::::::::::::::
 Db 236 SerGlyLysThrArgGluGlySerIleAlaAlaGluGluGlyValProLysLeuTyr 255
 ::::::::::::::
 QY 700 -----TTATCTACAAGTCTTTTACATTTAAGACAAAGTGGATGT 738
 ::::::::::::::
 Db 256 LeuValIleGlnLysGlnGluLeuValGlnAspLeuValLeuValAlaThrValGlyCys 275
 ::::::::::::::
 QY 739 TATGGAACTGGGATGGAGTGTGTTGATCCCGGATCTCTCAATAACAGCATCATCT 798
 ::::::::::::::
 Db 276 SerArgSerLeuSerPheGlu-----ProAspGlyGlnIleArgAlaSerSer 291
 ::::::::::::::
 QY 799 GTGCTGAGTGGACTGACACACAGGCAAGAACAGTTCGAAACCCCAAAAGCCAGG 858
 ::::::::::::::
 Db 292 SerTrpGlnSerValAsnGluSerGlyAspGlnValHisTrpSerProGlyGlnAlaArg 311
 ::::::::::::::
 QY 859 CTGAAAAAACCTGGACCCCTTGGCTGCTTTTCCACTGATGAA-----TAC 906
 ::::::::::::::
 Db 312 LeuGlnAspGlnGlyProSerTrpAlaSerGlyAspSerSerAsnAsnHisLysProArg 331
 ::::::::::::::
 QY 907 CAGTGGTTCAATAGATTGTAATAAGAAAGAAATAACAGGCAATTATACCACTGGA 966
 ::::::::::::::
 Db 332 GluTrpLeuGluIleAspLeuGlyGlyLysLysIleThrGlyIleArgThrGly 351
 ::::::::::::::
 QY 967 TCACCATGTTGGAGCACAATTACTATGTCTGCCCTACAACTCTCTACAGTGTATGAT 1026
 ::::::::::::::
 Db 352 SerThrGlnSerAsnPheAsnPheTyrValLysSerPheValMetAsnPheLysAsnAsn 371
 ::::::::::::::
 QY 1027 GGCAGAGAAAGTGGTGTACAGAGAGCGCTGCTGGGAGCAAGATAAGATATTTCAAGGA 1086
 ::::::::::::::
 Db 372 AsnSerLysTrpLysThrTyrLysGlyIleValAsnAsnGluGluLysVal---ArgGly 390
 ::::::::::::::
 QY 1087 AAC 1089


```
QY 976 GTGAGCACAATTACTATGTCTGCTACAGAACTCTGTACAGTGTATGATGGCAGAAA 1035
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 352 ThrGlnLysGlyTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1036 TGAAGTGTGTACAGAGAGCTGTGTGGAGCAGATATTAAGTATTTCAAGAAACAAGAT 1095
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysValPheGlnAlaAsnAsp 389
QY 1096 TATCACCAGGATGTGGTAACTATTTTGGCCACCAATATTCCACGTTTATTAGAGTG 1155
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 390 AlaThrGluLeuValLeuAsnTyrHisThrProLeuLeuThrArgPheIleArgIle 409
QY 1156 AATCTACCACTACAGGAGCAAAATTTGCCATGAAATGGAGTGTCTCGAGTGTACGTTT 1215
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1216 ATTCTAAAGTGTCTCTCCAAACTTACTCAACCTCCACTCCTCGGAACACCAATGAC 1275
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1276 CTCAAAACACTACAGCCCTCCAAAATAGCAAAAGGTGCTGCCCAAAATTT----- 1329
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTrp 460
QY 1330 -----ACGCAACCACTACAACCTCGCAGTAGCAATGAATTTCTCGCACAGACAGAA 1380
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgAsnProGln 480
QY 1381 CAACAACCTGCCAGTCTGTATCATCAGAAATPACTACCGTAACCTCCAAATGTAACCAAGAT 1440
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1441 GTA-----GCGTGGCTCCAGTTCTTCTGCTGCTGCTGCTGCTGCTGCTG 1473
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 501 ValIleIleGlnGlyAlaArgGlyAspSerIleThrAlaMetGluAlaArgAlaPhe 520
QY 1474 GTCATGTCTCTCACTACTCTCATTTCTCATATTAGTGTGCTGCTGCTGCTGCTGCTG 1524
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1525 -----AGAAACAGAAAGAAAGAACTGAAGGCACC-----TATGACTTACCTTACTGG 1572
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 ProArgThrGlnGlnProLysLeuPheGluGlyAsnMethIstYrAspThrPro----- 558
QY 1573 GACCGGCAGGTGGTGGAAAGGAATGAACAGTTTCTCTCGAAAGCAGTGGACCAT 1632
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 559 -----AspIleArgArgPhe----- 563
QY 1633 GAGGAACCCAGTTCGCTATAGCAGCAGCGAGTAAATACACCTGAGTCCCAAGAGAAGTC 1692
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 564 GluProValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1693 ACCACAGTGTGCGAGTGTCTGTCAGAGTAT-----GCTCAGCCACTGGTAGGAGGA 1746
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1747 ATTGTTGGTACACTTCATCAAGATCTACCTTAAACCA-----GAAGAAGGA 1794
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
QY 1795 AAGAAGCAGGCTAT-----GCAGACCTAGATCCTTCAACTCA 1833
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerG1 643
QY 1834 CCAGGGCAGGAAGTTTATCATGCTATGCTGAACCACTCCCAATACGGGCGCTGAGTAT 1893
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTrpMetTyrAspArgAlaLys 663
QY 1894 GCACCCCAATCATCATGACATGTCAGGGCACCCCA----- 1930
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 663 strPleuGlnSerThrTrpIleSerSerAlaAsnProAsnAspArgThrPheProAspAs 683
```

```
QY 1931 -----CAACTTCAG-----TTGGT 1944
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 683 pLysAsnPheLeuLysLeuGlnSerAspGlyGlyArgGluGlyGlnPheGlyArgLeuI1 703
QY 1945 CAGCCCTCCACATCCCACTTTCAAGGCTACGGGGAACCAACCTCCCACTAGTGGAACT 2004
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 703 eSerProPro-ValHisLeu-----ProArgSerProValCysMetGluP 718
QY 2005 TACAATACACTTCTCTC-----C 2022
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 718 heGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuGlnValValArgGluAlaArg 738
QY 2023 AGCACTGACAGCTGCTC-----CTCAGCCAGCCAGCCAGTATGATATCCCGAAAGCTGG 2076
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 738 lngluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlySerGluTrpLysHisGly 757
QY 2077 AGCCAGGTCTACTGCTCCCAAGAGGATTTGGTACCAGGTGCTACAGAGCACACAGAA 2136
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 758 ArgIleIleLeuProSerTyrAsp---MetGluTyrGlnIle-----ValPheGluGly 774
QY 2137 GTATCAGGACGAGGAGGATGGGAATGTGATGTT 2172
   :::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 775 valleGlyLysGlyArgSerGlyGluIleSerIle 786
```

Search completed: January 21, 2003, 09:35:35
Job time : 69.9883 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 09:21:35 ; Search time 37.4534 Seconds
(without alignments)
11242.484 Million cell updates/sec

Title: US-10-060-830-2
Perfect score: 3969
Sequence: 1 atgcctctgtttctctctact.....tttttaagaaaaatcctttga 2190

Scoring table:

	BLOSUM62	
	Xgapop 10.0 ,	Xgapext 0.5
	Ygapop 10.0 ,	Ygapext 0.5
	Fgapop 6.0 ,	Fgapext 7.0
	Delop 6.0 ,	Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

Command line parameters:

```

MODEL=framer_n2p.model -DEV=xlh
O/cgcn2_1/USPTO_spoof/US10060830/runat_16012003_091239_25859/app_query.fasta_1.6933
-D8-pr73 -USPTO_fastan -SUFFIX=n2p.tpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-QB-PR73 -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cd1 -LIST=45
-UNITS=bits -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -MAXLEN=2000000000
-OUTFWP=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830_@CGN1_1_1668runat_16012003_091239_25859 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGEV -TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

```
Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	383	9.6	927	1	KQ0948	A5 antigen precurs
2	361	9.1	2211	1	JR0905	coagulation factor
3	355.5	9.0	2183	2	T42764	coagulation factor
4	351	8.8	2224	1	KF005	coagulation factor
5	349	8.8	401	2	S65138	glycoprotein anti
6	349	8.8	427	2	S74211	PAS-6/7 protein p
7	343	8.6	218	2	A47285	milk fat globule p
8	340	8.6	409	2	T11743	pp47 protein - pig
9	339	8.5	427	2	JC4915	ags protein precu
10	321	8.1	463	1	A38479	milk fat-globule m
11	311.5	7.8	2133	2	T42763	coagulation factor
12	310	7.8	2319	2	A47004	coagulation factor
13	308.5	7.8	216	2	A44258	factor VIII-associ
14	308.5	7.8	2351	1	E2HD	coagulation factor

15	224.5	5.7	845	2	TJ5256
16	222.5	5.6	1524	2	TJ3037
17	219.5	5.5	3623	2	TJ08618
18	215.5	5.4	3133	2	T52093
19	211	5.3	3623	2	TQ9456
20	207.5	5.2	1283	2	TJ13799
21	204	5.1	719	2	S11739
22	200.5	5.1	855	2	S42621
23	198	5.0	277	2	A1735
24	198	5.0	707	2	TJ2218
25	195	4.9	823	1	A58788
26	194.5	4.9	730	1	BWHU1
27	193.5	4.9	1594	2	T30549
28	193	4.9	275	2	TJ6506
29	193	4.9	986	1	B58788
30	193	4.9	1035	1	A43090
31	192.5	4.9	991	2	I49540
32	190.5	4.8	694	2	TJ6554
33	190	4.8	1464	2	S58984
34	189.5	4.8	705	1	C1JHUB
35	188.5	4.7	1070	2	TJ1069
36	187	4.7	913	2	A48280
37	185.5	4.7	276	2	A47290
38	185.5	4.7	402	2	JH0403
39	185.5	4.7	688	1	C1JHUS
40	184	4.6	2083	2	T42721
41	183.5	4.6	449	2	A5362
42	183.5	4.6	910	2	A53137
43	183	4.6	2403	2	A59386
44	180	4.5	1057	1	A39288
45	175	4.4	1004	2	TJ30338
46	175	4.4	1004	2	TJ30338
47	175	4.4	1004	2	TJ30338
48	175	4.4	1004	2	TJ30338
49	175	4.4	1004	2	TJ30338
50	175	4.4	1004	2	TJ30338
51	175	4.4	1004	2	TJ30338
52	175	4.4	1004	2	TJ30338
53	175	4.4	1004	2	TJ30338
54	175	4.4	1004	2	TJ30338
55	175	4.4	1004	2	TJ30338
56	175	4.4	1004	2	TJ30338
57	175	4.4	1004	2	TJ30338
58	175	4.4	1004	2	TJ30338
59	175	4.4	1004	2	TJ30338
60	175	4.4	1004	2	TJ30338
61	175	4.4	1004	2	TJ30338
62	175	4.4	1004	2	TJ30338
63	175	4.4	1004	2	TJ30338
64	175	4.4	1004	2	TJ30338
65	175	4.4	1004	2	TJ30338
66	175	4.4	1004	2	TJ30338
67	175	4.4	1004	2	TJ30338
68	175	4.4	1004	2	TJ30338
69	175	4.4	1004	2	TJ30338
70	175	4.4	1004	2	TJ30338
71	175	4.4	1004	2	TJ30338
72	175	4.4	1004	2	TJ30338
73	175	4.4	1004	2	TJ30338
74	175	4.4	1004	2	TJ30338
75	175	4.4	1004	2	TJ30338
76	175	4.4	1004	2	TJ30338
77	175	4.4	1004	2	TJ30338
78	175	4.4	1004	2	TJ30338
79	175	4.4	1004	2	TJ30338
80	175	4.4	1004	2	TJ30338
81	175	4.4	1004	2	TJ30338
82	175	4.4	1004	2	TJ30338
83	175	4.4	1004	2	TJ30338
84	175	4.4	1004	2	TJ30338
85	175	4.4	1004	2	TJ30338
86	175	4.4	1004	2	TJ30338
87	175	4.4	1004	2	TJ30338
88	175	4.4	1004	2	TJ30338
89	175	4.4	1004	2	TJ30338
90	175	4.4	1004	2	TJ30338

ALIGNMENTS

RESULT 1

 JQ0948

 A5 antigen precursor - African clawed frog

 C:Species: Xenopus laevis (African clawed frog)

 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

 C:Accession: JH0466; JQ0948

 R:Author: S.; Hirata, T.; Agata, K.; Mochii, M.; Eguchi, G.; Fujisawa, H.

 Neuron 7, 295-307, 1991

 A:Title: The A5 antigen, a candidate for the neuronal recognition molecule, has homol

 A:Reference number: JH0466; MUID:9137458; PMID:1908252

 A:Accession: JH0466

 A:Molecule type: mRNA

 A:Residues: 1-927 <AK>

 A:Cross-references: GB:D010467; GB:D01077; NID:g222962; PIDN:BAA01360.1; PID:g222963

 A:Experimental source: tadpole, brain

 A:Note: this protein has motifs homologous to complement components C1r and C1s and t

 C:Comment: This protein is a neuronal cell surface molecule involved in the neuronal

 C:Superfamily: Xenopus A5 antigen; C1r/C1s repeat homology; discoidin I amino-termina

 C:Keywords: duplication; glycoprotein; transmembrane protein

 F:1-21/Domain: signal sequence #status predicted <SIG>

 F:22-927/Product: A5 antigen #status predicted <ASA>

 F:27-138/Domain: C1r/C1s repeat homology <C1R1>

 F:147-262/Domain: C1r/C1s repeat homology <C1R>

 F:274-424/Domain: discoidin I amino-terminal homology <DN1>

 F:430-584/Domain: discoidin I amino-terminal homology <DN2>

 F:646-812/Domain: MAM homology <MAM>

 F:861-883/Domain: transmembrane #status predicted <TM>

 F:150,261,300,523,844/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.:	1,18e-20	Length:	927
Score:	383.00	Matches:	138
Percent Similarity:	38.88%	Conservative:	70
Best Local Similarity:	25.79%	Mismatches:	197
Query Match:	9.65%	Indels:	130
DB:	1	Gaps:	22

US-10-060-830-2 (1-2190) x JQ0948 (1-927)

local

F:29-345/Domain: A1 <DA1>
F:33-329/Domain: ferroxidase repeat homology <FO1>
F:346-695/Domain: A2 <DA2>
F:351-688/Domain: ferroxidase repeat homology <FO2>
F:696-1564/Domain: B <DOB>
F:1175-1437/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F:1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>
F:1565-1892/Domain: A3 <DA3>
F:1572-1892/Domain: ferroxidase repeat homology <FO3>
F:1654-1752/Region: phospholipid binding #status predicted
F:1893-2051/Domain: C1 <DC1>
F:1893-2048/Domain: discolidin I amino-terminal homology <DNI>
F:2052-2211/Domain: C2 <DC2>
F:2052-2208/Domain: discolidin I amino-terminal homology <DN2>
F:167-239,248-329,499-525/Disulfide bonds: #status experimental
F:225-239,248-329,499-525/Disulfide bonds: 780,902,952,964,1044,1053,1062,1071,1078,1094,
F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:363-697/1537/Binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:533-534/Cleavage site: Arg-Gly (protein C) #status predicted
F:607-688,1712-1738,1894-2048,2053-2208/Disulfide bonds: #status predicted
F:741-742/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F:1034-1035/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F:1564-1565/Cleavage site: Arg-Ser (thrombin) #status experimental

Alignment Scores:
Pred. No.: 6,06e-19 Length: 2211
Score: 361.00 Matches: 79
Percent Similarity: 57.8% Conservative: 17
Best Local Similarity: 47.5% Mismatches: 60
Query Match: 9.40% Indels: 10
DB: 1 Gaps: 3

US-10-060-830-2 (1-2190) x KFE05 (1-2211)
QY 724 AGACAGTGGATGTTATGACACATGGGAGTGGTGTGTCGGGATCCTCAA 783
Db 2049 GluValAsnGlyCysSerThrProLeuGlyMetGluSerGlyLysLeuGln 2068
QY 784 ATACAGCATCATCTGCTGGAG-----TGGACTGACACACAGGGCAAGAACAGT 837
Db 2069 IleThrAlaSerSerPheLysLysSerTrp-----GlyAsnTyr 2082
QY 838 TGAACACCAAAAGCCAGGTGAAACCTGGA-----CCGCTTGGGCTGCTTT 891
Db 2083 TrpGluProPheLeuAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2102
QY 892 GCCACTGATGAATACCATGTTGACAAATAGATTGAAATGAAGAAAGAAATAACAGGC 951
Db 2103 AlaAsnAsnAsnGlnTrpLeuGlnIleAspLeuLysLysLysLysLysLysLysAla 2122
QY 952 ATTATACCATGTCATCCATGTTGGAGCACAATTACTATGTCGTGCTACAGAAATC 1011
Db 2123 IleValThrGlnGlyCysLysSerLeuSerSerGluMetTyrValLysSerTyrThrIle 2142
QY 1012 CTCTACAGTATGATGGCAGAAATAGACTGTGTACAGAGAGCTGTGTGGAGCAAGAT 1071
Db 2143 HisTyrSerAspGlnGlyThrAspTrpLysProTyrArgGluLysSerSerMetValAsp 2162
QY 1072 AGCATATTCAAGAAACAAAGATTATCACCAGATGTCGTAATACTTTTCCACCA 1131
Db 2163 LysIlePheGluGlyAsnAsnValArgGlyHisValLysAsnPheAsnProPro 2182
QY 1132 ATTATTGCACGTTTATTAGTGAATCTTACCATCATCCATGGCAGAAATTTGCCATGAA 1191
Db 2183 IleIleSerArgPheIleArgIleIleProLysThrTrpAsnGlnSerIleAlaLeuArg 2202
QY 1192 ATGAGCTGCTCGGATG 1209
Db 2203 LeuGluLeuPheGlyCys 2208

RESULT 3
T42764

coagulation factor V - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42764
R:Yang, T.L.; Cui, J.; Rehmtulla, A.; Yang, A.; Moussalli, M.; Kaufman, R.J.,
Blood 91, 4593-4599, 1998
A:Title: The structure and function of murine factor V and its inactivation by F
A:Reference number: 22270; PMID:9828202; PMID:9616155
A:Accession: T42764
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2183 <YAN>
A:Cross-references: EMBL:U52925; NID:g3219690; PID:g3219691; PIDN:AAC99553.1
C:Function:
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discolidin I amino-terminal homology; ferroxidase
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasm
F:350-682/Domain: ferroxidase repeat homology <FOX1>
F:1541-1864/Domain: ferroxidase repeat homology <FOX2>

Alignment Scores:
Pred. No.: 1.56e-18 Length: 2183
Score: 355.50 Matches: 79
Percent Similarity: 54.31% Conservative: 28
Best Local Similarity: 40.10% Mismatches: 75
Query Match: 8.96% Indels: 15
DB: 2 Gaps: 4

US-10-060-830-2 (1-2190) x T42764 (1-2183)
QY 631 GTAATTAGTAAGGTATTCCTATTATGAAGTCTTTGGCTAACACGTCACATCTGTG 690
Db 1995 IleValAlaArgTyrIleArgIleHisProThrLysSerTyrAsnArgProThr----- 2012
QY 691 GTGGGACACTTCTACAGTCTTTTACATTTAAGCAAGTGGATGTTATGGAACACTG 750
Db 2013 -----LeuArgLeuGluLeuGlnGlyCysGluValAsnGlyCysSerThrProLeu 2029
QY 751 GGATGGAGTCTGGTGTATCGGGATCCTCAATTAACAGCATCATCTGTGCTGGAG--- 807
Db 2030 GlyLeuGluAspGlyArgIleGlnAspLysGlnIleThrAlaSerPheLysSer 2049
QY 808 ---TGGACTGACACACAGGGCAAGAGAACAGTTGAAACCAAAAGACAGCTGAAA 864
Db 2050 TrpTrpGlyAspTyr-----TyrGluProSerLeuAlaArgLeuAsn 2063
QY 865 AAACCTGGA-----CCGCTTGGGCTGCTTTGCCACTGATGAATACCAAGTGTACAA 918
Db 2064 AlaGlnGlyArgValAsnAlaTrpGlnAlaLysAlaAsnAsnLysGlnTrpLeuGln 2083
QY 919 ATAGATTGAATAGGAAAGAAATAACAGGCATTATACCACTGGATCCACCATGGTG 978
Db 2084 ValAspLeuLeuLysLysLysValThrAlaIleValThrGlnGlyCysLysSerLeu 2103
QY 979 GACACATAATTACTATGTCGTGCTACAGATCTGTACAGTGTATGATGGCAGAAATGG 1038
Db 2104 SerSerGluMetTyrValLysSerTyrSerIleGlnTyrSerAspGlnGlyValAlaTrp 2123
QY 1039 ACTGTGTACAGAGACCTGCTGTGGAGCAAGATAAGATATTTCAAGAAACAAAGATTAT 1098
Db 2124 LysProTyrArgGlnLysSerSerMetValAspLysIlePheGluGlyAsnSerAsnThr 2143
QY 1099 CACAGGATCGGTATAAATCTTTTGGCCCAATATTGACGCTTTTATTAGAGTGAAT 1158
Db 2144 LysGlyHisMetLysAsnPheAsnProIleIleSerArgPheIleArgIleIle 2163
QY 1159 CTTACCCCAATGGCAGCAGAAATTCATGAAATGAGCTGCTCGGATG 1209
Db 2164 ProLysThrTrpAsnGlnSerIleAlaLeuArgLeuGluLeuPheGlyCys 2180

RESULT 4
KFE05
coagulation factor V precursor [validated] - human

N:Alternate names: coagulation labile factor; proaccelerin
 C:Species: Homo sapiens (man)
 C:Date: 19-May-1989 #sequence_revision 02-Jun-1995 #text_change 08-Dec-2000
 C:Accession: A56172; A42344; A28028; A27498; A25897
 R:Crripe, L.D.; Moore, K.D.; Kane, W.H.
 Biochemistry 31, 3777-3785, 1992
 A:Title: Structure of the gene for human coagulation factor V.
 A:Reference number: A42344; MUID:92232668; PMID:1567832
 A:Accession: A56172
 A:Molecule type: DNA
 A:Molecule type: mRNA
 A:Residues: 1-2224 <CR1>
 A:Cross-references: GB:J05368
 A:Accession: A42344
 A:Molecule type: DNA
 A:Residues: 48-58;79-89;120-130;191-201;239-249;313-323;368-378;428-437;461-471;533-542;
 2070;2111-2120;2172-2181 <CR2>
 R:Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kaufm
 Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
 A:Title: Complete cDNA and derived amino acid sequence of human factor V.
 A:Reference number: A28028; MUID:87260886; PMID:3110773
 A:Accession: A28028
 A:Molecule type: mRNA
 A:Residues: 1-857; R', 859-864, R', 866-924, 'E', 926-1763, 'I', 1765-2212, 'T', 2214-2224 <JEN>
 A:Cross-references: GB:M16967
 A:Note: parts of this sequence, including the amino end of the mature protein, were dete
 R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.
 Biochemistry 26, 6508-6514, 1987
 A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum
 A:Reference number: A27498; MUID:88107560; PMID:2827731
 A:Accession: A27498
 A:Molecule type: mRNA
 A:Residues: 1-1284; 'I', 1286-1600 <KAN>
 A:Cross-references: GB:M17785
 A:Note: parts of this sequence were determined by protein sequencing
 R:Kane, W.H.; Davie, E.W.
 Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
 A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homolog
 A:Reference number: A25897; MUID:86313665; PMID:3092220
 A:Accession: A25897
 A:Molecule type: mRNA
 A:Residues: 1188-1215,1315-2224 <KA2>
 A:Cross-references: GB:M14335
 A:Note: parts of this sequence were determined by protein sequencing
 R:Keller, F.G.; Ortel, T.L.; Quinn-Alten, M.A.; Kane, W.H.
 Biochemistry 34, 4118-4124, 1995
 A:Title: Thrombin-catalyzed activation of recombinant human factor V.
 A:Reference number: A56133; MUID:95210278; PMID:7696276
 A:Contents: annotation; thrombin cleavage sites
 C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
 C:Genetics:
 A:Gene: GDB:F5
 A:Cross-references: GDB:1119896; OMIM:227400
 A:Map position: 1q23-1q23
 A:Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65
 C:function:
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
 C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:29-224/Product: coagulation factor V #status predicted <MAT>
 F:29-737/Product: coagulation factor Va heavy chain #status experimental <MAT>
 F:29-345/Domain: A1 <DA1>
 F:33-329/Domain: ferroxidase repeat homology <FO1>
 F:346-691/Domain: A2 <DA2>
 F:351-684/Domain: ferroxidase repeat homology <FO2>
 F:692-1573/Domain: B <DOB>
 F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
 F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
 F:1574-1905/Domain: A3 <DA3>
 F:1581-1905/Domain: ferroxidase repeat homology <FO3>
 F:1667-1765/Region: phospholipid binding #status predicted
 906-2064/Domain: C1 <DC1>

F:1906-2061/Domain: discoidin I amino-terminal homology <DN1>
 F:2065-2224/Domain: C2 <DC2>
 F:2065-2221/Domain: discoidin I amino-terminal homology <DN2>
 F:51,55,239,297,460,468,554,741,752,760,776,782,821,938,977,1074,1083,1103,1106,1479,
 F:167-193,248-329,500-526,603-684,1725-1751,1907-2061,2066-2221/Dsulfide bonds: #sta
 F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
 F:363,693,1546/Binding site: sulfate (Tyr) (covalent) #status predicted
 F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
 F:382,1338/Binding site: carboxydrate (Asn) (covalent) #status predicted
 F:534-535/Cleavage site: Arg-Gly (protein C) #status absent
 F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experiment
 F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experime
 F:1573-1574/Cleavage site: Arg-Ser (Chromobin) #status experimental

Alignment Scores: 3,42e-18 Length: 2224
 Pred. No.: 351.00 Matches: 77
 Percent: 59.04% Conservatives: 21
 Best Local Similarity: 46.39% Mismatches: 58
 Query Match: 8.84% Indels: 10
 DB: 1 Gaps: 3

US-10-060-830-2 (1-2190) x KFHU5 (1-2224)

QY	724	AGACAAAGTGGATGTTATGGAACTGGGATGGAGTCTGGTGTGATCGCGGATCTCAA	783
DB	2062	GluValAsnGlyCysSerThrProLeuGlyMetGluAsnGlyLysIleGluAsnLysGln	2081
QY	784	ATAACAGCATCATCTCTGCTGGAG-----TGCATGACACACACAGGCGGCAAGAACAGT	837
DB	2082	IleThrAlaSerSerPheLysLysSerTrpLysAspTyr-----	2095
QY	838	TGGAACCCCAAAAGCCAGGCTGAAACCTGGA-----CCGCCTTGGCTGCTTTT	891
DB	2096	TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys	2115
QY	892	GCACGTGATGAATACACAGTGGTTTACAAATAGATTGAATAGGAAGAAATAACAGGC	951
DB	2116	AlaAsnAsnLysGlnTrpLeuGluIleAspLeuLysIleLysIleThrAla	2135
QY	952	ATTATAACCACTGATCCACCATGGTGGAGCAAAATTAATCTATGCTGCTGCTACAGATC	1011
DB	2136	IleIleThrGlnGlyCysLysSerLeuSerSerGluMetTrpValLysSerTrpThrIle	2155
QY	1012	CTGTACAGTATGATGGCGAGAAATGGACTGTGTACAGAGAGCGCTGTGTGGAGCAAGAT	1071
DB	2156	HisTyrSerGluGlnGlyValGluTrpLysProTyrArgLeuLysSerSerMetValasp	2175
QY	1072	AGATATTTCAGGAAACAAAGATTATCACAGGATGTGGTAACTATTAATCTTTGCCACCA	1131
DB	2176	LysIlePheGluGlyAsnThrAsnThrLysGlyHisValLysAsnPheAsnProPro	2195
QY	1132	ATTATTGACAGTTTATTAGATGAATCTACCATGGCAGCAGAAATTCGCATGAAA	1191
DB	2196	IleIleSerArgPheIleArgValIleProLysThrTrpAsnGlnSerIleAlaLeuArg	2215
QY	1192	ATFGAGCTGCTCGGATGT	1209
DB	2216	LeuGluLeuPheGlyCys	2221

RESULT 5
 S65138
 glycoprotein antigen MGp57/53, mammary gland - bovine (fragment)
 N:Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
 C:Accession: S65138; G48394
 R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
 Biochim. Biophys. Acta 1245, 385-391, 1995
 A:Title: Molecular cloning of glycoprotein antigens MGp57/53 recognized by monoclonal
 A:Reference number: S65138; MUID:96125736; PMID:8541316
 A:Accession: S65138
 A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-401 <AOK>

R:Wather, I.H.; Banghart, L.R.; Lane, W.S.

Biochem. Mol. Biol. Int. 29, 545-554, 1993

A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig II-like sequences.

A:Reference number: A48394; MUID:93250576; PMID:8485470

A:Accession: G48394

A>Status: preliminary

A:Molecule type: protein

A:Residues: 207-220 <MAT>

A:Experimental source: milk

A:Note: sequence extracted from NCBI backbone (NCBIP:131457)

C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homology

C:Keywords: glycoprotein

F:1-32/Domain: EGF homology (fragment) <EG1>

F:40-79/Domain: EGF homology <EG2>

F:82-239/Domain: discoidin I amino-terminal homology <DN1>

F:243-401/Domain: discoidin I amino-terminal homology <DN2>

Alignment Scores:

Pred. No.:	3,7e-18	Length:	401
Score:	349.00	Matches:	77
Percent Similarity:	56.59%	Conservative:	26
Best Local Similarity:	42.31%	Mismatches:	68
Query Match:	8.79%	Indels:	11
DB:	2	Gaps:	5

US-10-060-830-2 (1-2190) x S65138 (1-401)

QY	684	ATCTGT-----GGTGGGACACTTATCTACAGTCTTTTACATTTAAGCAAGTGGATG	737
Db	225	IleCysHisArgGlyCysThrLeuArgPheGlu-LeuLeuGlyCysGluLeuAsnGlyCy	244
QY	738	TTATGGACACTGGGATGGAGTCTGGTGTGTCGCGGATCCTCAATAAAGCATCATC	797
Db	244	sThrGluProLeuGlyLeuLysAspAsnThrIleProAsnLysGlnIleThrAlaSerSe	264
QY	798	TGTGCTGGAG---TGGACTGACACACAGG---CAAGAGACACTGGGAACCCAAAAA	851
Db	264	rTyTyThrTrp-----GlyLeuSerAlaPheSerTrpPheProTyTy	280
QY	852	AGCAGGCTGAAAAAACCTGGACCG-----CCTTGGGCTGCTTTTGGCCACTGATGA	905
Db	280	rAlaArgLeuAspAsnGlnGlyLysPheAsnAlaTrpThrAlaGlnThrAsnSerAlaSe	300
QY	906	CCAGTGGTTACAATAGATTGTAAGAAAGAAATAACAGGCATTATAACCACTGG	965
Db	300	rGluTrpLeuGlnIleAspLeuGlySerGlnLysArgValThrGlyIleThrGlnG	320
QY	966	ATCCACCATGGTGAGCACAAATTAATGTCTGCTGCCTACAGATCCTGTACAGTGATGA	1025
Db	320	yAlaArgAspPheGlyHisIleGlnTyrValAlaAlaTyrArgValAlaTyrGlyAspAs	340
QY	1026	TGGCAGAAATGGACTGTGACAGAGAGCTGGTGGAGCAAGATAAGATATTTCAGG	1085
Db	340	pGlyValThrTrpThrGluTyrLysAspProGlyAlaSerGluSerLysIlePheProG	360
QY	1086	AAACAAGATATACACAGGATGTCGTAATACTTTTGGCCACCAATATTATGGACGTTT	1145
Db	360	yAsnMetAspAsnSerHisLysLysAsnIlePheGluThrProPheGlnAlaArgPh	380
QY	1146	TATTAGATGATCTTACCCTATGGCAGCAAAATTTGCCATGAATGGAGCTGCTCGG	1205
Db	380	eValArgIleGlnProValAlaTrpHisAsnArgIleThrLeuArgValGluLeuGln	400
QY	1206	ATGT 1209	
Db	400	yCys 401	

RESULT 6

S74211

PAS-6/7 protein precursor - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000

C:Accession: S74211; S78114; S24181

R:Hvarregard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.

Eur. J. Biochem. 240, 628-636, 1996

A:Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk I II-like sequences.

A:Reference number: S74211; MUID:97008954; PMID:8856064

A:Accession: S74211

A:Molecule type: mRNA

A:Residues: 1-427 <HVA>

A:Cross-references: EMBL:X91895; NID:g1632778; PIDN:CAA62997.1; PID:g1632779

A:Accession: S78114

A:Molecule type: protein

A:Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-

R:Kim, D.H.; Kanno, C.; Mizokami, Y.

Biochim. Biophys. Acta 1122, 203-211, 1992

A:Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, f

A:Reference number: S23926; MUID:92353107; PMID:1643094

A:Accession: S24181

A:Molecule type: protein

A:Residues: 383-394 <KIM>

C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom

C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-427/Product: PAS-6/7 protein #status experimental <MAT>

F:24-58/Domain: EGF homology <EG1>

F:66-105/Domain: EGF homology <EG2>

F:108-265/Domain: discoidin I amino-terminal homology <DN1>

F:269-427/Domain: discoidin I amino-terminal homology <DN2>

F:24-35;29-47;49-58;66-77;71-94;96-105/Disulfide bonds: #status predicted

F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental

F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental

F:59-227/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:109-265;252-256;270-427/Disulfide bonds: #status experimental

Alignment Scores:

Pred. No.:	3,73e-18	Length:	427
Score:	349.00	Matches:	77
Percent Similarity:	56.59%	Conservative:	26
Best Local Similarity:	42.31%	Mismatches:	68
Query Match:	8.79%	Indels:	11
DB:	2	Gaps:	5

US-10-060-830-2 (1-2190) x S74211 (1-427)

QY	584	ATCTGT-----GGTGGGACACTTATCTACAGTCTTTTACATTTAAGCAAGTGGATG	737
Db	251	IleCysHisArgGlyCysThrLeuArgPheGlu-LeuLeuGlyCysGluLeuAsnGlyCy	270
QY	738	TTATGGACACTGGGATGGAGTCTGGTGTGTCGCGGATCCTCAATAAAGCATCATC	797
Db	270	sThrGluProLeuGlyLeuLysAspAsnThrIleProAsnLysGlnIleThrAlaSerSe	290
QY	798	TGTGCTGGAG---TGGACTGACACACAGG---CAAGAGACACTGGGAACCCAAAAA	851
Db	290	rTyTyThrTrp-----GlyLeuSerAlaPheSerTrpPheProTyTy	306
QY	852	AGCAGGCTGAAAAAACCTGGACCG-----CCTTGGGCTGCTTTTGGCCACTGATGA	905
Db	306	rAlaArgLeuAspAsnGlnGlyLysPheAsnAlaTrpThrAlaGlnThrAsnSerAlaSe	326
QY	906	CCAGTGGTTACAATAGATTGTAAGAAAGAAATAACAGGCATTATAACCACTGG	965
Db	326	rGluTrpLeuGlnIleAspLeuGlySerGlnLysArgValThrGlyIleThrGlnG	346
QY	966	ATCCACCATGGTGAGCACAAATTAATGTCTGCTGCCTACAGATCCTGTACAGTGATGA	1025
Db	346	yAlaArgAspPheGlyHisIleGlnTyrValAlaAlaTyrArgValAlaTyrGlyAspAs	366
QY	1026	TGGCAGAAATGGACTGTGACAGAGAGCTGGTGGAGCAAGATAAGATATTTCAGG	1085
Db	366	pGlyValThrTrpThrGluTyrLysAspProGlyAlaSerGluSerLysIlePheProG	386

QY 1086 AACAAGATTATCACAGGATGCGTAACTATTTTCCACCAATATTATTCACGTTT 1145
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 386 YasnMetAspAsnAsnSerHisLysLysAsnIlePheGluThrProPheGlnAlaArgPh 406
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 1146 TATTAGAGTGAATCTACCAATGAGCAGCAGAGAAATGGCATGAAATGGAGCTGCTCGG 1205
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 406 eValArgIleGlnProValAlaTrpHisAsnArgIleThrLeuArgValGluLeuGlu 426
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 1206 ATGT 1209
|||||
Db 426 YCys 427
|||||
RESULT 7
A47285
milk fat globule protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
C:Accession: A47285
R:La Rocca, D.; Peterson, J.A.; Ureia, R.; Kuniyoshi, J.; Bistrain, A.M.; Ceriani, R.L.
Cancer Res. 51, 4994-4998, 1991
A:Title: A M-r 46,000 human milk fat globule protein that is highly expressed in human B
A:Reference number: A47285; MUID:91371351; PMID:1909932
A:Accession: A47285
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <LAR>
A:Cross-references: GB:S56151; NID:q235396; PIDN:AAB19771.1; PID:q235397
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
F:1-56/Domain: discoidin I amino-terminal homology (fragment) <DN1>
F:60-218/Domain: discoidin I amino-terminal homology <DN2>
Alignment Scores:
Pred. No.: 9,49e-18 Length: 218
Score: 343.00 Matches: 71
Percent Similarity: 57.23% Conservative: 28
Best Local Similarity: 41.04% Mismatches: 68
Query Match: 8.64% Indels: 6
DB: 2 Gaps: 3
US-10-060-830-2 (1-2190) x A47285 (1-218)
QY 700 TTATCTACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACTGGGATGGAG 759
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 49 LeuArgPheGluLeuLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 68
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 760 TCIGGTGTGATCGGATCTCAATAACAGCATCATCTGCTGGAG---TGGACTGAC 816
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 69 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerSerTyrLysThrTrpGlyLeu 88
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 817 CACACAGGGCAAGAACAGTGTGGAACCCCAAGGAGGCTGGAACACCTGGA--- 873
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 89 HisLeu-----PheSerTrpAsnProSerTyrAlaArgLeuAspLysGlnGlyAsn 105
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 874 ---CCGCTTGGCTGCTTTTGGCAGTGTGATACACAGTGGTTTACAAATAGATTGTAAT 930
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 106 PheAsnAlaTrpValAlaGlySerTyrGlyAsnAspGlnTrpLeuGlnValAspLeuGly 125
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 931 AAGGAAAGAAATAACAGCAATATACCACTGGATCCACCATGTTGGAGCACAATTAC 990
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 126 SerSerLysGluValThrGlyIleIleThrGlnGlyAlaArgAsnPheGlySerValGln 145
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 991 TATGTGCTGCCACAGAACTCTGACAGTGTGATGGCAGAAATGAGCTGTGTACAGA 1050
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 146 PheValAlaSerTyrLysValAlaTyrSerAsnAspSerAlaAsnTrpThrGluTyrGln 165
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 1051 GAGCCTGGTGGGCAAGATAAGATATTTCAAGGAACAAGATTATCACAGGATGTG 1110
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 166 AspProArgThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisSerHisLys 185
||||| ||| : : : ||||| ||| ||| ||||| |||
1111 CGTAAATACCTTTTGGCACCACCAATATTGACGTTTATTATGAGTGAATCATCCCAATGG 1170
||||| ||| : : : ||||| ||| ||| ||||| |||
186 LysAsnLeuPheGluThrProIleLeuAlaArgTyrValArgIleLeuProValAlaTrp 205
||||| ||| : : : ||||| ||| ||| ||||| |||

QY 1171 CAGCAGAAATGCCATGAAATGGAGTCTCGGATGT 1209
: : : ||||| : : : : : ||||| ||||| ||||| |||||
Db 206 HisAsnArgIleAlaLeuArgGluLeuLeuGlyCys 218
||||| ||| : : : ||||| ||| ||| ||||| |||
RESULT 8
T11743
pp47 protein - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
R:Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toe
Biol. Reprod. 58, 1057-1064, 1998
A:Title: Molecular cloning and characterization of P47, a novel boar sperm-associated
A:Reference number: Z17325; MUID:98206817; PMID:9546740
A:Accession: T11743
A:Status: preliminary; translated from CB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-409 <ENS>
A:Cross-references: EMBL:X11683; NID:g2652927; PIDN:CAA72379.1; PID:g2652928
A:Experimental source: testis
C:Function:
A:Description: may be involved in membrane remodeling and/or function as a zona pellu
C:Superfamily: unassigned EGF-related proteins; EGF homology
F:6-40/Domain: EGF homology <EGF>
Alignment Scores:
Pred. No.: 1,76e-17 Length: 409
Score: 340.00 Matches: 76
Percent Similarity: 56.59% Conservative: 27
Best Local Similarity: 41.76% Mismatches: 68
Query Match: 8.57% Indels: 11
DB: 2 Gaps: 5
US-10-060-830-2 (1-2190) x T11743 (1-409)
QY 684 ATCTGT-----GGTGGACACTTATCTACAAGTCTTTTACATTTAAGACAAGTGGATG 737
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 233 IleCysHisArgGlyCysThrLeuArgPheGlu-LeuLeuGlyCysGluLeuSerGlyCy 252
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 738 TTATGGAACACTGGGATGGAGTCTGTGTGATCGCGGATCCTCAATAACAGCATCATC 797
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 252 salagiUpLeuGlyLeuLysAspAsnThrIleProAsnLysGlnIleThrAlaSerSe 272
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 798 TGTGCTGGAG---TGGACTCACCACACAGG---CAAGAGAACAGTGGAAACCAAAAA 851
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 272 rPheTyrArgThrTrp-----GlyLeuSerAlaPheSerTyrTrpPheTy 288
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 852 AGCCAGCTGAAACACCTGGACCG-----CCTTGGCTGCTTTTGGCAGCTGATGAATA 905
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 288 rAlaArgLeuAspAsnGlnGlyLysPheAsnAlaTrpThrAlaGlnSerAsnSerAlaSe 308
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 906 CCAGTGGTTACAATAGATTGTAATAGGAAAGAAATACAGCAGCATTAACCACTGG 965
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 308 rGluTrpLeuGlnIleAspLeuGlySerGlnArgValThrGlyIleIleThrGlnGl 328
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 966 ATCCACCATGGTGGAGCACAATTACTATGTCTGCCTACAGAATCCTGTACAGTATGA 1025
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 328 yAlaArgAspPheGlyHisIleGlnTyrValAlaIleAlaTyrLysValAlaTyrSerAspAs 348
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 1026 TGGGCAAGAAATGGAGTGTGTACAGAGAGCCCTGTGTGGAGCAAGATAGATATTTCAAGG 1085
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 348 pGlyValSerTrpThrGluTyrArgAspGlnGlyAlaLeuGluGlyLysIlePheProGl 368
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 1086 AAACAAGATTATCACAGGATGTCCTAATAACTTTTGGCCACCAATATTGACGTTT 1145
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 368 yAsnLeuAspAsnAsnSerHisLysLysAsnMetPheGluThrProPheLeuThrArgPh 388
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 1146 TATTAGAGTGAATCTACCAATGGCAGCAGAAATGGCAATGAGGAGTCTCGG 1205
||||| ||| : : : ||||| ||| ||| ||||| |||
Db 388 eValArgIleLeuProValAlaTrpHisAsnArgIleThrLeuArgValGluLeuGlu 408
||||| ||| : : : ||||| ||| ||| ||||| |||
QY 1206 ATGT 1209
|||||

Db 408 yCys 409

RESULT 9

ags protein precursor - rat
 N;Alternate names: O-acetyl-Gd3 ganglioside
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
 C;Accession: JG4915
 R;Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
 Biochem. Biophys. Res. Commun. 225, 932-938, 1996
 A;Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
 A;Reference number: JG4915; MUID:96374422; PMID:8780713
 A;Accession: JG4915
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-427 <OGU>
 A;Cross-references: DBJB:DR84058; NID:g1620006; PIDN:BNAL2210.1; PID:g1620007
 A;Experimental source: CST cell
 C;Comment: This protein is required for the O-acetylation of disialoganglioside sialic
 C;Genetics:
 A;Gene: ags
 C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homoloq
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;28-60/Domain: EGF homology <EG1>
 F;68-107/Domain: EGF homology <EG2>
 F;110-267/Domain: discoidin I amino-terminal homology <DN1>
 F;271-427/Domain: discoidin I amino-terminal homology <DN2>

Alignment Scores:

Alignment Scores:			
Pred. NO.:	2, file-17	Length:	427
Score:	339.00	Matches:	128
Percent Similarity:	38.66%	Conservative:	56
Best Local Similarity:	26.89%	Mismatches:	160
Query Match:	8.54%	Indels:	132
DB:	2	Gaps:	19

US-10-060-830-2 (1-2190) x JC4915 (1-427)

QY	13	CTCTGCTCTTACTTGTCCCTGCTCTGCTGCTGAGGACCTGGAGCCCGACGCAAGGTGAT	72
Db	7	LeuAlaAlaLeuCysGlyValLeuLeuCysAlaSerGlyLeuPheAlaAlaSerGlyAsp	26
QY	73	GGATGT-----GGACACACTGTACTAGGC-----	96
Db	27	PheCysAspSerSerLeuCysLeuAsnGlyGlyThrCysLeuMetGlyGlnAspAsnAsp	46
QY	97	-----CTGAGAGTGGAAACCTTACATTCATTAACCTACCACAGACACCTAT	141
Db	47	IleIyrCysLeuCysProGluGlyPheThrGlyLeuValCysAsnGluThrGluGlyGly	66
QY	142	CCCAACAGC---ACTGTTTGTGAATGGGAGATCCGTGTAAGACTGGAGAGAGAGTCCGC	198
Db	67	ProCysSerProAsnProCysPheHisAspAlaLysCysLeuValThrGluAspThr	85
QY	199	ATCAAAATTGGTGACTTTGACATTGAAGATCTGATCTTGTGCACCTTAAATTACTTGAGA	258
Db	86	---GlnArgGlyAspIlePheThrGlyIleCysGlnCysProValGlyTyrSerGly	104
QY	259	ATTTATAATGCAATTGGAGTCAGCAGAACCTGAAATAGGC-----	297
Db	105	IleHisCysGluLeuGlyCysSer---ThrLysLeuGlyLeuGluGlyGlyAlaIleAla	123
QY	298	-----AAATACTGTGGT---CTGGGGTTGCAATGAACCAT	330
Db	124	AspSerGlnIleSerAlaSerSerValIyrMetGlyPheMetGlyLeuGln-----	140
QY	331	TCAATTGAATCAAAAGCAATGAATCACATTCGTGTTTCATGAGTGGCATC-----	381
Db	141	-----ArgTrpGlyProGluLeuAlaArgLeuIyrArgThrGlyIleValAsnAla	157
QY	381	-----	381

RESULT 10

RESULT TO
A36479

A36479
milk fat globule membrane protein - mouse

C:Species: Mus musculus (house mouse)

C; Species: Mus musc
C; Date: 10-Sep-1999

C; Date: 10-Sep-1993
C; Accession: A36479

C; ACCESSION: A30475
R: Stubbs, J. D.: Lek

R; Stubbs, J. D.; Lex
proc. Natl. Acad. Sci.

Proc. Natl. Acad. S
A:Title: cDNA clonin

A;Title: cDNA clone1
A:Reference number:

A;Reference number:
A:Accession: A36479

A;Accession: A36479

Db 1906 rSerLeuAspGlyArgAsnTrpGlnSerTyrArgGlyAsnSerThrGlyThrLeuMetVa 1926
QY 661 -----AGTTCTTTGGCTACACAGTC----- 681
Db 1926 lPhePheGlyAsnValAspAlaSerGlyIleLysHisAsnIlePheAsnProProIleVa 1946
QY 682 -----ACATCTGTGGTGGGACACATCTTACAAAG 710
Db 1946 lAlaArgTyrIleArgLeuHisProThrHisTyrSerIleArgSerThrLeuArgMetG1 1966
QY 711 TCTTTTACATTTAAGACAAGTGGATGTTATGAACACATCGGGATGGAGTGGTGGTAT 770
Db 1966 uLeuMetGlyCysAspLeuAsnSerCysSerMetProLeuGlyMetGlnAsnLysAlaI1 1986
QY 771 CGGGATCTCAATTAACACATCATCTGTCTGGAGTGGACTGACCACACAGGGCAAGA 830
Db 1986 eSerAspSerGlnIleThrAlaSerHisLeuSerAsnIlePheAlaThr----- 2003
QY 831 GAACAGTTGAAACCCCAAAAGCCAGCTGAAAAAACCTGGA-----CCGCCCTTGGGC 884
Db 2004 -----TrpSerProSerGlnAlaArgLeuHisLeuGlnGlyArgThrAsnAlaIatPAr 2021
QY 885 TGCTTTTGCACATGATTAACACAGTGGTGTACAAATAGATTTGAATAAGGAAAAAGAAAT 944
Db 2021 gProArgValSerSerAlaGluGluTrpLeuGlnValAspLeuGlnLysThrValLysVa 2041
QY 945 AACAGGATTAATACCACTGGATCCACCATGGTGGGACACAAATTAATGCTGTGCTA 1004
Db 2041 lThrGlyIleThrThrGlnGlyValLysSerLeuLeuSerSerMetTyrValLysGluPh 2061
QY 1005 CAGATCTGTACATGATGATGGCAGAAATGGACGTGTGTACAGAGAGCCTGTGTGGA 1064
Db 2061 eLeuValSerSerSerGlnAspGlyArgArgTrpThrLeuPheLeuGlnAspGly----- 2079
QY 1065 GCAAGATAAGATTTTCAAGGAACAAAGATTAATCACCAGGATGCGGTAAATATCTTTT 1124
Db 2080 -HisThrLysValPheGlnGlyAsnGlnAspSerSerThrProValValAsnAlaLeuAs 2099
QY 1125 GCGACCAATTAATCCAGGTTTATAGATGAAATCCATACCAATGGCAGCAAAATGTC 1184
Db 2099 pProProLeuPheThrArgTyrLeuArgIleHisProThrSerTrpAlaGlnHisIleAl 2119
QY 1185 CATGAAATGGAGTGTCTCGATGTCAG 1212
Db 2119 aLeuArgLeuGluValLeuGlyCysGlu 2128
RESULT 12
A47004
coagulation factor VIII precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 18-Jun-1999
C:Accession: A47004
R:Elder, B.; Laktich, D.; Gitschler, J.
Genomics 16, 374-379, 1993
A:Title: Sequence of the murine factor VIII.cDNA.
A:Reference number: A47004; MUID:93300511; PMID:8314577
A:Accession: A47004
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2319 <ELD>
A:Cross-references: GB:I05573; NID:g192456; PIDN:AAA37385.1; PID:g192457
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:23-349/Domain: ferroxidase repeat homology <FOL>
F:402-730/Domain: ferroxidase repeat homology <FOL>
F:1686-2006/Domain: ferroxidase repeat homology <FOL>
F:2007-2156/Domain: discoidin I amino-terminal homology <DN1>
F:2160-2313/Domain: discoidin I amino-terminal homology <DN2>

Alignment Scores: 4.14e-15 Length: 2319
Pred. No.: 310.00 Matches: 96
Score:

Percent Similarity: 44.76% Conservative: 45
Best Local Similarity: 30.48% Mismatches: 113
Query Match: 7.81% Indels: 61
DB: 2 Gaps: 11
US-10-060-830-2 (1-2190) x A47004 (1-2319)
QY 306 TGGTCTGGGTTGCAATGAACCATTAATGAATCAAAAGGCAATGAATACATGCT 365
Db 2049 TrpSer-----ThrLysGluProPheSerTrpIle-LysValAspLeuAlaProMe 2066
QY 366 GTTCATGAGTGGAAATCCATGTTTCTGGA-----CG 395
Db 2066 tIleValHisGlyIleLysThrGlnGlyAlaArgGlnLysPheSerSerLeuTyrIleSe 2086
QY 396 CGGATTTTGGCCCTCATACCTCTGTATAGATAAACAAGATCTAATTAATCTTGTGGCAC 455
Db 2086 rGlnPheIleIleMetTyrSerLeuAspGlyLysLys-----TrpLeuSerTy 2102
QY 456 TGCATCCAAATTTTGGAACTGAGTTCAGTAACTAGTCCCGCAGCTGGTGTCTGCTTCC 515
Db 2102 rGlnGlyAsn-----SerThrGlyThrLeuMetVa 2112
QY 516 TTTTGTGAGATATCTCGAACAATTCCTCATGGATATAGATTCCTCGCCATTGTGCAT 575
Db 2112 lPhe-----PheGlyAsnValAspSerSer----- 2120
QY 576 GGCTGGTGTGCATGTCAGAGTAGTGTCAAAACAGCTGGCGGCAAACTAGTGTGTAAT 635
Db 2121 -----GlyIleLysHisAsnSerPheAsnProProIle----- 2132
QY 636 TAGTAAAGGATTTCCCTTATTATGAAGTCTTTGGCTTAACAACGTCACATCTGTGGTGGG 695
Db 2132 eAlaArgTyrIleArgLeuHisProThrHis-----SerSerIleArgSe 2147
QY 696 ACATTTATACAGTCTTTTACATTTAAGACAAGTGGATGTTATGAACACTGGGAT 755
Db 2147 rThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerIleProLeuGlyMe 2167
QY 756 GGACTGTGTGTGTCGGGATCCTCAAAATACACATCATCTGTCTGTGGAGTGGACTGA 815
Db 2167 tGluSerLysValIleSerAspThrGlnIleThrAlaSerSerTyrPhe----- 2183
QY 816 CCACACAGGGCAAGACAGACAGTGTGGAACCCCAAAAGCCAGGCTGAAAAACCTGGA-- 873
Db 2184 ---ThrAsnMetPheAlaThrTrpSerProSerGlnAlaArgLeuHisLeuGlnGlyAr 2202
QY 874 ---CCGCCCTTGGGCTGCTTTTGGCACTGATGAATACCAGTGGTTACAAATAGATTTGAA 929
Db 2202 gThrAsnAlaTrpArgProGlnValAsnAspProLysGlnTrpLeuGlnValAspLeuG1 2222
QY 930 TAAGGAAAAAGAAATAACAGGCATTAATACCACTGGATCCACCATGGTGGAGCACAAATTA 989
Db 2222 nLysThrMetLysValThrGlyIleIleThrGlnGlyValLysSerLeuPheThrSerMe 2242
QY 990 CTATGTGTCTGCTTACAGAAATCCCTGATGATGATGGGAGCAAAATGAGCTGTGTACAG 1049
Db 2242 tPheValLysGluPheLeuIleSerSerSerGlnAspGlyHisIstTrpThrGlnIleLe 2262
QY 1050 AGAGCCTGTGTGGGAGCAAGATATATTTCAAGAAACAAAGATTTATCACCAGGATGT 1109
Db 2262 uTyrAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSerSerThrProMe 2280
QY 1110 GCGTAATAACTTTTGGCCCAATTAATTCACAGTGTATTTATAGAGTGAATCTTACCAATG 1169
Db 2280 tMetAsnSerLeuAspProProLeuLeuThrArgTyrLeuArgIleHisProGlnIleTr 2300
QY 1170 GCAGCAGAAAAATGGCATGAAATGGAGCTGCTCGGATGTCTAC 1212
Db 2300 pGluHisGlnIleAlaLeuArgLeuGluIleLeuGlyCysGlu 2314
RESULT 13
A44258

factor VIII-associated gene B hypothetical protein - human

C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C:Accession: A44258

R:Levinson, B.; Kenrick, S.; Gamel, P.; Fisher, K.; Gitschler, J.

Genomics 14, 585-589, 1992

A:Title: Evidence for a third transcript from the human factor VIII gene.

A:Reference number: A44258; MUID:93052386; PMID:1427887

A:Accession: A44258

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-216 <LEV>

A:Cross-references: GB:M90707; NID:g182316; PIDN:AAA58466.1; PID:g182317

C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase

F:1-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>

F:57-210/Domain: discoidin I amino-terminal homology <DN2>

Alignment Scores:

Pred. No.:	3,71e-15	Length:	216
Score:	308.50	Matches:	66
Percent Similarity:	54.49%	Conservative:	31
Best Local Similarity:	37.08%	Mismatches:	72
Query Match:	7.77%	Indels:	9
DB:	2	Gaps:	3

US-10-060-830-2 (1-2190) x A44258 (1-216)

QY	685	TCGTGGTGGGACATTATCTACAGTCTTTTACATTTAGACAAAGTGGATGTATGGA	744
Db	41		
QY	745	ACACGGGGATGGTGGTGGTGGATCGGATCGCTCAATAACAGCATCATCTGTGCTG	804
Db	61		
QY	805	GAGTGGACTGACCCACAGCGGCAAGAACAGTGGAAACCCAAACCCAGCCAGCTGAAA	864
Db	81		
QY	865	AAACCTGGA-----CCGCTTGGCTGCTTTTGCACATGATACCATCGGTTACAA	918
Db	96		
QY	919	ATGATTTGAATTAAGAAAGAAATAACAGCATTAACACCTGGATGATGATGATGATG	978
Db	116		
QY	979	GAGCAAAATTAATGTGCTGCTACAGAAATCCTGTACAGTATGATGGCAGAAATGG	1038
Db	136		
QY	1039	ACTGTGTACAGAGACCTGGTGGGACCAAGATAGATATTTCAGGAAACAAAGATTAT	1098
Db	156		
QY	1099	CACAGGATGCGTGAATAACATTTTGGCCACCAATTATTCACGCTTTTATAGATGAAT	1158
Db	174		
QY	1159	CTACCCATGCGCAGACAAATTCCTGATGAAATGGAGCTGCTGGATGTCAG	1212
Db	194		
RESULT 14			
EBHU			
coagulation factor VIII precursor [validated] - human			
N:Alternate names: antihemophilic factor A; coagulation factor VIIIC; procoagulant compo			
C:Species: Homo sapiens (man)			
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 08-Dec-2000			
C:Accession: I54318; A00525; I58059; A23584; A26174; A43986; S63527; S66445; B42			
R:Gitschler, J.; Wood, W.I.			
Hum. Mol. Genet. 1, 199-200, 1992			
A:Title: Sequence of the exon-containing regions of the human factor VIII gene.			

A:Reference number: I54318; MUID:93265012; PMID:1303178
A:Accession: I54318
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1921, 'S', 1923-2351 <RES>
A:Cross-references: GB:M88648; NID:g182381; PIDN:AAA52420.1; PID:g182383
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschler, J.; Keyt, B.; See
Nature 312, 330-337, 1984
A:Title: Expression of active human factor VIII from recombinant DNA clones.
A:Reference number: A00525; MUID:85061548; PMID:6438526
A:Accession: A00525
A:Molecule type: mRNA
A:Residues: 1-2351 <WOO>
A:Cross-references: EMBL:X01166; EMBL:X01179
R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Suitzman, L.A.; Buecker, J.L.; Pittman, D.
S, D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A:Reference number: I58059; MUID:85061550; PMID:6438528
A:Accession: I58059
A>Status: nucleic acid sequence not shown; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-74, 'V', 76-1259, 'E', 1261-2351 <RE2>
A:Cross-references: GB:K01740; NID:g182802; PIDN:AAA52484.1; PID:g182803
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.;
B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.
DNA 4, 333-349, 1985
A:Title: Characterization of the polypeptide composition of human factor VIII:C and
A:Reference number: A23584; MUID:86081164; PMID:3935400
A:Accession: A23584
A:Molecule type: mRNA
A:Residues: 1-2351 <TRU>
A:Cross-references: GB:M14113; NID:g182817; PIDN:AAA52485.1; PID:g182818
R:Eaton, D.; Rodriguez, H.; Vénhar, G.A.
Biochemistry 25, 505-512, 1986
A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleava
ity.
A:Reference number: A26174; MUID:86159740; PMID:3082357
A:Accession: A26174
A:Molecule type: protein
A:Residues: 20-367392-399, 'X', 401-402; 1668-1678; 1709-1722, 'D', 1723-1725; 1741-1755 <E>
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.
Biochemistry 31, 3315-3325, 1992
A:Title: Identification and functional importance of tyrosine sulfate residues with
A:Reference number: A42348; MUID:92207952; PMID:1554716
A:Accession: A42348
A:Molecule type: protein
A:Residues: 20-367356-371; 392-408; 582-594; 1668-1669, 'X', 1671; 1672-1692; 1693-1708; 170
A:Experimental source: recombinant material from Chinese hamster ovary cells
A:Note: sequence extracted from NCBI backbone and corrected to correspond with the p
E:Pay, P.J.; Smudzin, T.M.
J. Biol. Chem. 264, 14005-14010, 1989
A:Title: Intersubunit fluorescence energy transfer in human factor VIII.
A:Reference number: A43986; MUID:89340500; PMID:2503509
A:Accession: A43986
A:Molecule type: protein
A:Residues: 'X', 517-523; 1853-1860, 'X', 1862-1864, 'X', 1866 <PAY>
R:Leyte, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens,
J. Biol. Chem. 266, 740-746, 1991
A:Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential
A:Reference number: A56109; MUID:91093266; PMID:1898735
A:Contents: annotation; sulfation
R:Gitschler, J.; Wood, W.I.; Goralka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vén
Nature 312, 326-330, 1984
A:Title: Characterization of the human factor VIII gene.
A:Reference number: A56196; MUID:85061547; PMID:6438525
A:Contents: annotation; introns
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
Protein Sci. 4, 740-746, 1995
A:Title: Locations of disulfide bonds and free cysteines in the heavy and light chai
A:Reference number: A56216; MUID:95338127; PMID:7613471
A:Contents: annotation; disulfide bonds
A:Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls

R;Kjalke, M.; Heding, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
Eur. J. Biochem. 234, 773-779, 1995

A;Title: Amino acid residues 721-729 are required for full factor VIII activity.

A;Reference number: S63527; MUID:96163459; PMID:8575434

A;Accession: S63527

A;Molecule type: protein

A;Residues: 733-752;753-759 <KJA>

R;Lind, P.; Larsson, K.; Spira, J.; Sydow-Baekman, M.; Almstedt, A.; Gray, E.; Sandberg

Eur. J. Biochem. 232, 19-27, 1995

A;Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction

A;Reference number: S66445; MUID:96048024; PMID:7556150

A;Accession: S66445

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1668-1685 <LIN>

C;Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure ph

C;Genetics:

A;Gene: GDB:F8C

A;Cross-references: GDB:119124; OMIM:306700

A;Map position: Xq28-Xq28

A;Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63

C;Function:

A;Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro

A;Pathway: blood coagulation

C;Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase

C;Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla

F;1-19/Domain: signal sequence #status experimental <SIG>

F;20-2351/Product: coagulation factor VIII #status experimental <MAT>

F;20-740/Product: coagulation factor VIIIA heavy chain #status experimental <ACH>

F;23-348/Domain: A1 <DA1>

F;392-759/Domain: A2 <DA2>

F;402-730/Domain: ferroxidase repeat homology <FO1>

F;402-730/Domain: ferroxidase repeat homology <FO2>

F;1668-2351/Product: coagulation factor VIIa light chain #status experimental <ACL>

F;1709-2038/Domain: A3 <DA3>

F;1716-2038/Domain: ferroxidase repeat homology <FO3>

F;2039-2191/Domain: C1 <DC1>

F;2039-2188/Domain: discoidin I amino-terminal homology <DN1>

F;2192-2351/Domain: C2 <DC2>

F;2192-2345/Domain: discoidin I amino-terminal homology <DN2>

F;60-258,601,776,803,847,919,962,1020,1024,1074,1085,1204,1274,1278,1301,1319,1403,1

F;172-198,267-348,547-573,649-730,1851-1877,1918-1922,2040-2188/Disulfide bonds: #status

F;355-356/Cleavage site: Arg-Net (coagulation factor Xa, protein C) #status predicted

F;365,737,738,742,1683,1695/Binding site: sulfate (Tyr) (covalent) #status experimental

F;391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental

F;414,426/Binding site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental

F;759-760/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental

F;1667-1668/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental

F;1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental

F;1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental

F;2193-2345/Disulfide bonds: #status predicted

Alignment Scores:

Pred. No.: 5,38e-15 Length: 2351

Score: 308.50 Matches: 66

Percent Similarity: 54.49% Conservative: 31

Best Local Similarity: 37.08% Mismatches: 72

Query Match: 7.77% Indels: 9

DB: 3 Gaps: 3

US-10-060-830-2 (1-2190) x EZHU (1-2351)

Qy 685 TCTGTGGGACACTTATCTTACAGTCTTTTATCAATTAAGACAGTGGATGTTATGGA 744

Db 2176 SerileArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet 2195

Qy 745 ACATGGGATGGATCTGGTGTGATCCCGGATCCTCAAAATACAGCATCTCTGCTG 804

Db 2196 ProLeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyrPhe 2215

Qy 805 GAGTGGATGCCACACAGGCGACAGACAGTTCGAAACCCAAAAGCCAGGCTGAA 864

Db 2216 -----ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis 2230

Qy 865 AAACCTGGA-----CGCCTTGGGTGCTTTTGGCCACTCATCAATACAGTGGTTACAA 918

Db 2231 LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGln 2250

Qy 919 ATAGATTGAATAAGAAAGAAATAACAGGCAATTAACCACTGGATCCACCATGGTG 978

Db 2251 ValAspPheGlnLysThrMetLysValThrGlyValThrThrGlnGlyValLysSerLeu 2270

Qy 979 GAGCAAAATACATATGTGTGCTGCTCAGCAATCTGTACAGTATGATGGCAGAAATGG 1038

Db 2271 LeuThrSerMetTyrValLysGluPheLeuIleSerSerGlnAspGlyHisGlnTrp 2290

Qy 1039 ACTGTGTACAGAGAGCTGTGGAGCAAGATAAGATATTCAAGAAACAAGATTAT 1098

Db 2291 ThrLeuPheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer 2308

Qy 1099 CACCAGGATGTGGTGAATAACTTTTCCACCAATATTATTCACGCTTTTATTAGAGTAAT 1158

Db 2309 PheThrProValValAsnSerLeuAspProLeuLeuThrArgTyrLeuArgIleHis 2328

Qy 1159 CTTACCAATGGCAGCAGAAATTCCTCAAAATGGAGCTGCTCGATGTCAG 1212

Db 2329 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 2346

RESULT 15

JC5256

adipocyte transcription factor, AEBP1 - human

C;Species: Homo sapiens (man)

C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C;Accession: JC5256

R;Ohno, I.; Hashimoto, J.; Shimizu, K.; Takaoka, K.; Ochi, T.; Matsubara, K.; Okubo,

Biochem. Biophys. Res. Commun. 228, 411-414, 1996

A;Title: A cDNA cloning of human AEBP1 from primary cultured osteoblasts and its expr

A;Reference number: JC5256; MUID:97079196; PMID:8920928

A;Accession: JC5256

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-845 <OHN>

A;Cross-references: DDBJ:D86479; NID:g1468942; PIDN:BAAL3094.1; PID:g1468943

Alignment Scores:

Pred. No.: 9,38e-09 Length: 845

Score: 224.50 Matches: 52

Percent Similarity: 47.90% Conservative: 28

Best Local Similarity: 31.14% Mismatches: 68

Query Match: 5.66% Indels: 19

DB: 2 Gaps: 4

US-10-060-830-2 (1-2190) x JC5256 (1-845)

Qy 748 CTGGGATGGAGTGTGGTGTGATCGGGATCTCAATAACAGCATCATCTGTCTGGAG 807

Db 75 IleGlyMetGluSerHisArgIleGluAspAsnGlnIleArgAlaSerSerMetLeu--- 93

Qy 808 TGGACTGACCACACAGGCGACAGACAGTTCGAAACCCAAAAGCCAGGCTGAAAAA 867

Db 94 -----ArgHisGlyLeuGlyAlaGlnArgGlyArgLeuAsnMet 106

Qy 868 CCT-----GGACCGCTTGGCTGCTTTTGGCCACTGAT 900

Db 107 GlnThrGlyAlaThrGluAspAspTyrTyrAspGlyAlaTrpCysAlaGluAspAla 126

Qy 901 GAATACCACTGGTTACAAATAGATTGAATAGGAAAGAAAATAACAGGATTAATACC 960

Db 127 ArgThrGlnTrpIleGluValAspThrArgThrThrArgPheThrGlyValIleThr 146

Qy 961 ACTGGATCCCACTGGTGGAGCACAATTAATCTATGCTGCTGCCTACAGATCTCTACAGT 1020

Db 147 GlnGlyArgAspSerSerIleHisAspPheValThrPhePheValGlyPheSer 166

Qy 1021 GATGATGGCACAATGGACTGTGTACAGAGCGCTGTGGAGCGCTGTAAGATAATTT 1080

Search completed: January 21, 2003, 09:41:36
Job time : 55.4534 secs

Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluThrLeuTrp 460
Qy 1330 -----ACGCAACCACTACAACTCGCAGTAGCATGAATTCCTCGCAGACAGAA 1380
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyThrProArgAsnProGln 480
Qy 1381 CAACAACACTGCGCTGATGATCAGAAATACCTACCGTAATCCAAATGTACCAAGAT 1440
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
Qy 1441 GTA-----CCGCTGGCTGCAGTCTTGTCTCCCTGCTGCTG 1473
Db 501 ValIleleGlnGlyAlaArgGlyGlyAspSerIleThrAlaMetGluAlaArgAlaPhe 520
Qy 1474 GTCATGCTCCTCACTACTCTCATCTCATATAGTGTGCTGCTGGCACTGG----- 1524
Db 521 ValArgLysPheLysValSerLysSerLeuAsnGlyLysAspTrpGluThrIleGlnAsp 540
Qy 1525 -----AGAAACAGAAAGAAACAACTGAAGCACC-----TATGACTTACTTACTGG 1572
Db 541 ProArgThrGlnGlnProLysLeuPheGluGlyAsnMetHisTyraThrPro----- 558
Qy 1573 GACCGGCGAGTGTGGTGAAGGATGAAGCAGTTCTTCTCGTCAAAAGCAGTGACCAT 1632
Db 559 -----AspIleArgArgPhe----- 563
Qy 1633 GAGGAACCCAGTTCGCTATAGCAGCAGCGAAGTAAATACCTGAGTCCAGAGAAAGTC 1692
Db 564 GluProValProAlaGlnTyrrValArgValTyrrProGluArgTrpSerProAlaGlyIle 583
Qy 1693 ACCACAGTGTGCGGTGACTCTGCAGAGTAT-----GTCAGCAGCTGAGGAGGA 1746
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
Qy 1747 ATTGTGTGTACACTCATCAAGATCTACCTTAAACCA-----GAAGAAGA 1794
Db 604 LeuGlyProThrValLysSerGluGluThrThrProTyrrProMetAspGluAla 623
Qy 1795 AAAGAAGCAGGCTAT-----GCAGACCTAGATCTTACAACTCA 1833
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerG1 643
Qy 1834 CAGGGCAGGAGTTATATGCTATGCTGAAACACTCCCAATGACGGGCGCTGAGT 1893
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTrpMetTyraArgAla 663
Qy 1894 GCAACCCCAATCATCATGATGATGTCAGGCGCACCCCA----- 1930
Db 663 sTrpLeuGlnSerThrTrpIleSerSerAlaAsnProAsnAspArgThrPheProAspAs 683
Qy 1931 -----CAACTTCAG-----TTGCT 1944
Db 683 PlysAsnPheLeuLysLeuGlnSerAspGlyArgGluGlyGlnPheGlyArgLeuI1 703
Qy 1945 CAGCCCTCCATCATCTTCAAGCTACGGGGAACCACTCCCTCCCTACTAGTGGAACT 2004
Db 703 eSerProPro-ValHisLeu-----ProArgSerProValCysMetGluP 718
Qy 2005 TACAATACACTCTCTC----- 2022
Db 718 heGlnTyrrGlnAlaMetGlyHisGlyValAlaLeuGlnValValArgGluAlaArgG 738
Qy 2023 AGGCTGACAGCTGCTC-----CFCAGCCAGGCGCCAGTATGATACCCGAAAGCTGGG 2076
Db 738 InGluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlySerGluTrpLysHisGly 757
Qy 2077 AAGCCAGCTCTACCTGCCCCAGCAAGTGTGTACCAAGTGTCCACAGCAGCACAGAA 2136
Db 758 ArgIleIleLeuProSerTyraAsp---MetGluTyrrGlnIle-----ValPheGluGly 774
Qy 2137 GTATCAGGACGAGGAGGATGGGAATGTATGTT 2172
Db 775 ValIleGlyLysGlyArgSerGlyGluIleSerIle 786

RESULT 2

NRP2_MOUSE STANDARD; PRT; 931 AA.
ID NRP2_MOUSE 035373; 035374; 035376; 035377; 035378;
AC 035375; 035373; 035374; 035376; 035377; 035378;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
DE NRP2.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=BALB/C;
RX MEDLINE=97470888; PubMed=9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
affinity receptor for the semaphorins Sema E and Sema IV but not Sema
III.";
RT III.";
RL Neuron 19:547-559(1997).
CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC -1- SUBUNIT: NEUROPILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
NEUROPILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN E RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS: A22 (SHOWN HERE), A0, A5, A17,
BO AND B5; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING CNS, PNS AND IN SOME
NONNEURAL TISSUES INCLUDING LIMB BUDS, DEVELOPING BONES, MUSCLES,
INTESTINAL EPITHELIUM, KIDNEY, LUNG AND SUBMANDIBULAR GLAND.
CC -1- DEVELOPMENTAL STAGE: THE EXPRESSION PATTERN IS VERY DYNAMIC AND
IS DEVELOPMENTALLY REGULATED.
CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF022856; AAC53379.1; -
CC EMBL; AF022854; AAC53377.1; -
CC EMBL; AF022855; AAC53378.1; -
CC EMBL; AF022857; AAC53380.1; -
CC EMBL; AF022858; AAC53381.1; -
CC EMBL; AF022861; AAC53382.1; -
CC HSSP; P12259; 1CZT.
CC MGD; MGI:1100492; Nrp2.
CC InterPro; IPR000859; CUB_domain.
CC InterPro; IPR000421; FA58_C.
CC InterPro; IPR000998; MAM_domain.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00629; MAM; 1.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00042; CUB; 2.
CC SMART; SM00231; FA58C; 2.
CC SMART; SM00137; MAM; 1.
CC PROSITE; PS01180; CUB; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PROSITE; PS0060; MAM_2; 1.
KW Transmembrane; Glycoprotein; Neutrone; Signal; Repeat; Receptor;
KW Alternative splicing.
FT SIGNAL 1 POTENTIAL.

[illegible]

Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTrp 460
QY 1330 -----ACCAACCACTACAACTCGCAGTAGCAATGTAATTCCTCGCAGACAGAA 1380
Db 461 SerProSerAlaAlaArgValSerArgSerGlyTyrPheProArgAsnProGln 480
QY 1381 CAACAACCTGCCAGTCCTGATACAGAAATACCTACCGTAACCTCAATACCAAGAT 1440
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1441 GTA-----GGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1473
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1474 GTCATGGCTCCTACACACCTCCTATTCATATATAGTGTGTGCTGCTGCTGCTGCTG 1524
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpLutLutIleGlnAsp 540
QY 1525 -----AGAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1572
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1573 GACCGGGCAGGTGTGGGAAAGGATGAAGCAGTTCTTCTCGCAAAAGCAGTGGACCAT 1632
Db 559 -----AspIleArgArgPheAspPro----- 565
QY 1633 GAGAAACCCAGTTCGTATAGCAGCAGCAGGAAATTAATACCTGAGTCCACAGAAAGTC 1692
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTyrProAlaGlyIle 583
QY 1693 ACCACAGTCTCCAGCTGACTCTCAGAGTAT-----GCTCAGCCACTGCTAGGAGA 1746
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1747 ATGTGTGTACACTTCATCAAGATCTACCTTTAAACCA-----GAAGAAGA 1794
Db 604 LeuGlyProThrValLysSerGluGluThrThrProTyrProMetAspGluAspAla 623
QY 1795 AAAGAACGAGGTAT-----GCAGACTAGATCTTACACTCA 1833
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspAspLysAspLeuGln-LeuProSerG1 643
QY 1834 CCAGGCGAGGAGTTATCATGCTC-----ATGCTGAACACTCCCA 1875
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTyrValTyrAspHisAla 663
QY 1876 A---TTACGGCCCTGAGTATGCAACCCCAATCATCATCATCATCATCATCATCATCAT 1930
Db 663 strLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 1931 -----CAACTCAG----- 1939
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgG1 696
QY 1940 -----TTGGTCAGCCCTCCACATCCACTTTCAGGCTACGGGACCA 1983
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 1984 CTTCCCTCCCTAGTGGGAACCTTACATACCTCTCTCT----- 2021
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2022 -----CAGGACTGACACTGCTC-----CTCAGCCCGCCGCG 2055
Db 731 lnValValArgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2056 TATGATACCCGAAAGCTGGGAGCAGCTCTACCTGCTGCTGCTGCTGCTGCTGCTGCTG 2115
Db 751 GlySerGluTrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2116 GCGCCACAGCAGCACAAAGATATCATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2163
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu 783

RESULT 3

NR1P1_HUMAN STANDARD; PRT; 923 AA.
AC 014786; 060461;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuropilin-1 precursor (Vascular endothelial cell growth factor 165 receptor).
GN NR1P1 OR NRP OR VEGF165R.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM).
RX MEDLINE=97433084; PubMed=9288753;
RA He Z., Tessier-Lavigne M.;
RT "Neuropilin-1 is a receptor for the axonal chemorepellent semaphorin III.";
RL Cell 90:739-751(1997).
RN [2]
RP SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM), AND SEQUENCE OF 22-39.
RC TISSUE=Brain;
RX MEDLINE=98188099; PubMed=9529250;
RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745(1998).
RN [3]
RP SEQUENCE FROM N.A. (SOLUBLE/SNRP1 ISOFORM), AND SEQUENCE OF 22-31.
RC TISSUE=Prostatic adenocarcinoma;
RX MEDLINE=20183929; PubMed=10688880;
RA Gagnon M.L., Bielenberg D.R., Gerechtman Z., Miao H.-Q., Takashima S., Soker S., Klagsbrun M.;
RT "Identification of a natural soluble neuropilin-1 that binds vascular endothelial growth factor: In vivo expression and antitumor activity.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:2573-2578(2000).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid form of vascular endothelial growth factor (VEGF) and of placenta growth factor-2, but only neuropilin-2 functions as a receptor for the 145-amino acid form of VEGF.";
RL J. Biol. Chem. 275:18040-18045(2000).
CC -!- FUNCTION: THE MEMBRANE-BOUND ISOFORM IS A RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF CERTAIN NEURONAL CIRCUITS AND IN ORGANOREGULATION OUTSIDE THE NERVOUS SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PCF, THE VEGF-165 ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS. IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS.
CC -!- FUNCTION: THE SOLUBLE/SNRP1 ISOFORM BINDS VEGF-165 AND APPEARS TO INHIBIT ITS BINDING TO CELLS. IT MAY ALSO INDUCE APOPTOSIS BY SEQUESTERING VEGF-165. MAY BIND AS WELL VARIOUS MEMBERS OF THE SEMAPHORIN FAMILY. ITS EXPRESSION HAS AN AVERSE EFFECT ON BLOOD VESSEL NUMBER AND INTEGRITY.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. THE SOLUBLE/SNRP1 ISOFORM IS SECRETED.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; MEMBRANE-BOUND (SHOWN HERE) AND SOLUBLE/SNRP1; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: THE EXPRESSION OF MEMBRANE-BOUND AND SOLUBLE/SNRP1 ISOFORMS DOES NOT SEEM TO OVERLAP. THE MEMBRANE-BOUND ISOFORM IS EXPRESSED BY THE BLOOD VESSELS OF DIFFERENT TISSUES. IN THE DEVELOPING EMBRYO IT IS FOUND PREDOMINANTLY IN THE NERVOUS SYSTEM. IN ADULT TISSUES, IT IS HIGHLY EXPRESSED IN HEART AND PLACENTA; MODERATELY IN LUNG, LIVER, SKELETAL MUSCLE, KIDNEY

CC	AND PANCREAS; AND LOW IN ADULT BRAIN. THE SOLUBLE/SNRP1 ISOFORM IS
CC	FOUND IN LIVER HEPATOCYTES, KIDNEY DISTAL AND PROXIMAL TUBULES.
CC	-1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC	-1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC	-1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AF018956; AAC51759.1; -;
DR	EMBL; AF016050; AAC12921.1; -;
DR	EMBL; AF145712; AAF44344.1; -;
DR	HSSP; P12259; ICZT.
DR	Genew; HGNC:8004; NRPI.
DR	MIM; 602069; -;
DR	InterPro; IPR000859; CUB_domain.
DR	InterPro; IPR000421; FA5B_C.
DR	InterPro; IPR001092; HLH_Basic.
DR	InterPro; IPR000998; MAM_domain.
DR	Pfam; PF00431; CUB; 2.
DR	Pfam; PF00629; MAM; 1.
DR	Pfam; PF00754; F5_F8_type_C; 2.
DR	PRINTS; PR00020; MAMDOMAIN.
DR	SMART; SM00042; CUB; 2.
DR	SMART; SM00231; FA58C; 2.
DR	SMART; SM00137; MAM; 1.
DR	PROSITE; PS01180; CUB; 2.
DR	PROSITE; PS01285; FA58C_1; 2.
DR	PROSITE; PS01286; FA58C_2; 2.
DR	PROSITE; PS00740; MAM_1; 1.
DR	PROSITE; PSS0060; MAM_2; 1.
KW	Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW	Alternative splicing.
FT	SIGNAL 1 21
FT	CHAIN 22 923
FT	DOMAIN 22 856
FT	TRANSMEM 857 879
FT	DOMAIN 880 923
FT	DOMAIN 27 141
FT	DOMAIN 147 265
FT	DOMAIN 275 424
FT	DOMAIN 431 583
FT	DOMAIN 645 811
FT	DISULFID 27 54
FT	DISULFID 82 104
FT	DISULFID 147 173
FT	DISULFID 206 228
FT	DISULFID 275 424
FT	DISULFID 431 583
FT	CARBOHYD 150 150
FT	CARBOHYD 261 261
FT	CARBOHYD 300 300
FT	CARBOHYD 522 522
FT	CARBOHYD 842 842
FT	VARSPLIC 642 644
FT	VARSPLIC 645 923
FT	CONFLICT 26 26
FT	CONFLICT 749 749
FT	CONFLICT 855 855
FT	SEQUENCE 923 AA; 103120 MW; ADEADC4A849ESD57 CRC64;
SQ	
CC	-----
CC	Alignment Scores:
CC	Pred. No.: 2.04e-21
CC	Score: 400.50
CC	Percent Similarity: 42.33%
CC	Best Local Similarity: 28.08%
CC	Query Match: 10.09%
CC	Length: 923
CC	Matches: 130
CC	Conservative: 66
CC	Mismatches: 178
CC	Indels: 89

Db 679 rPheProAspArgSerPheLeuArgLeuGlnSerAspSerGlnArgGluGlyGlnTy 699
 Qy 1934 -CTTCAAGTTGTCAGCCCTCCACATCCACTTTCAAGGCTAGGGGAACCAACCTCCCCCA 1992
 Db 699 rAlaArgLeuLeuSerProPro-ValHisLeu-----ProArgSerProV 714
 Qy 1993 CTAGTGGGAATTACATACACTCTCTCTC----- 2021
 Db 714 alCysMetGluPheGlnTyGlnAlaThrGlyArgGlyValalaleuGlnValVala 734
 Qy 2022 -----CAGACTGACAGCTGCTC-----CTCAGCCAGCCAGCTATGATACC 2064
 Db 734 rGluAlaSerGlnCysSerTyLeuLeuTrpValile-ArgGluAspGlnGlyGlyGlu 753
 Qy 2065 CCGAAGCTGGGAAGCCAGGCTCTACCTGCCAGACGAATTGGTTACCAAGTGCACAC 2124
 Db 754 TrpLysHisGlyArgIleleuLeuProSerTyRasp---MetGluTyRGlNile----- 770
 Qy 2125 AGCACACAAGATATCATCAGGACGAGGAGGATGGGAA 2163
 Db 771 ValPheGluGlyValileGlyLysGlyArgSerGlyGlu 783
 RESULT 5
 NRPI_MOUSE
 ID NRPI_MOUSE STANDARD; PRT; 923 AA.
 AC P97333;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neuropilin-1 precursor (A5 protein).
 GN NRPI OR NRP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN=BALB/c; TISSUE=Embryonic brain;
 RX MEDLINE=96353149; PubMed=8748368;
 RA Kawakami A., Katsukawa T., Takagi S., Fujisawa H.;
 RT "Developmentally regulated expression of a cell surface protein,
 RT neuropilin, in the mouse nervous system.";
 RL J. Neurobiol. 29:1-17(1996).
 CC -!- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
 CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT
 CC BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGF, THE VEGF-165
 CC ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN
 CC INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS.
 CC IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: NERVOUS SYSTEM.
 CC -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D50086; BAA08789.1; -
 DR HSP; P12259; 1C2T.
 DR MGD; MGI:106206; Nrp.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR000998; MAM_domain.

DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00629; MAM; 1.
 DR Pfam; PF00734; F5_F8_type_C; 2.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; SM00137; MAM; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00740; MAM_1; 1.
 DR PROSITE; PS00660; MAM_2; 1.
 KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor.
 FT SIGNAL 1 21
 FT CHAIN 22 923
 FT DOMAIN 22 856
 FT TRANSMEM 857 879
 FT DOMAIN 880 923
 FT DOMAIN 27 141
 FT DOMAIN 147 265
 FT DOMAIN 275 424
 FT DOMAIN 431 583
 FT DOMAIN 645 811
 FT DISULFID 27 54
 FT DISULFID 82 104
 FT DISULFID 147 173
 FT DISULFID 206 228
 FT DISULFID 275 424
 FT DISULFID 431 583
 FT CARBOHYD 150 150
 FT CARBOHYD 261 261
 FT CARBOHYD 300 300
 FT CARBOHYD 522 522
 FT CARBOHYD 842 842
 SQ SEQUENCE 923 AA; 103020 MW; 0644B8A170796808 CRC64;
 Alignment Scores:
 Pred. No.: 9,18e-21 Length: 923
 Score: 391.50 Matches: 129
 Percent Similarity: 40.04% Conservative: 60
 Best Local Similarity: 27.33% Mismatches: 176
 Query Match: 9.86% Indels: 107
 Db: 1 Gaps: 18
 US-10-060-830-2 (1-2190) x NRPI_MOUSE (1-923)
 Qy 16 CTGCTCTTACTGTCTCTCTCTCTGCTGCTGAGGACGCTGAGCCAGCAAGGTGATGGA 75
 Db 7 LeuLeuCysAlaThrLeuAlaLeuAlaLeuAlaLeuAlaGlyAlaPheArgSerAspLys 26
 Qy 76 TGTGACACACACTGTACTAGGCCCTGAGAGTGGAACTTACATCCATAAATACCCACAG 135
 Db 27 CysGlyGlyThrIleLysIleGluAsnProGlyTyLeuThrSerProGlyTyProHis 46
 Qy 136 ACCTATCCACACAGCAGCTGTTGTAATGGAGATCCCTGTA---AAGATGGAGAGAGA 192
 Db 47 SerTyHisProSerGluLysCysGluTrpLeuIleGluAlaProGluProTyRglNarg 66
 Qy 193 GTTCCATCAAAATTTGGT---GACTTTCATTTGAGATTTCTGATTTCTGTGACATTTAAT 249
 Db 67 IleIleIleAsnPheAsnProHisPheAspLeuGluAspArgAsp---CysLysTyRasp 85
 Qy 250 TACTTGAGAAATTTAATGAATTTGAGTTCAGCAGCAACTGAAATAGCAATATCTGTGGT 309
 Db 86 TyrValGluValIleAspGlyGluAsnGluGlyGlyArgLeuTrpGlyLysPheCysGly 105
 Qy 310 CTGGGGTTCGAAATGAACCATTC---ATTGAATCAAAGGCAATGAATCATCATGCTGCTG 366
 Db 106 -----LysIleAlaProSerProValSerSerGlyProPheLeuPheIleLys 122
 Qy 367 TTCATGAGTGGAAATCCATGTTCTGACGCGGATTTTGGCCCTCATCTACTCTGTTATAGAT 426
 Db 123 PheValSerAspTyRgluThrHisGlyAlaGlyPheSerIleArgTyRgluIlePheLys 142

QY 427 AAACAAGATCTAATTACTTGTGGACACTGCATCCCAATTTTGGACACTGAGTTCAGT 486
Db 143 ArgelyPro-----GlucySerGlnAsnYrThrAlaPro----- 154
QY 487 AAGTACTGCCAGCTGCTGCTGCTTTCCTGAGATATCTGGAACAATTCCTCAT 546
Db 155 -----ThrGlyValIleLysSer 160
QY 547 ---GGATATAGAGATCTCCGCCATCTGCGATG-----GCTGGT 592
Db 161 ProGlyPheProGluLysYrProAsnCysLeuGluCysThrYrIleLysPheAlaPro 180
QY 583 GTGCATGACGAGTACTG-----AGTGTGTATAGTAAGGATATCC----- 651
Db 181 LysMetSerGluIleLeuGluPheGluSerPheAspLeuGluIleAspSerAsnPro 200
QY 610 TTGGCGGCCAATC-----AGTGTGTATAGTAAGGATATCC----- 651
Db 201 ProGlyGlyMetPheCysArgYrAspArgLeuGluIleTyrAspGlyPheProGluVal 220
QY 652 -----TATTATGAAGTCTCTTGGCTAACACAGCTCATCTGTGGTG 593
Db 221 GlyProHisIleGlyArgYrCysGlyGlnLysThrProGlyArgIleArgSerSer 240
QY 694 GGACACTTACTACAAGTCTTTTACA----- 720
Db 241 GlyValLeuSerMetValPheYrThrAspSerAlaIleAlaLysGluGlyPheSerAla 260
QY 721 -----TTTAAGCAAGTGGATGTTATGGA 744
Db 261 AsnYrSerValLeuGlnSerSerIleSerGluAspPheLys-----CysMetGlu 277
QY 745 ACACCTGGGATGAGTCTGCTGCTGGGATCTCTCAATACACAGCATCATCTGTGCTG 804
Db 278 AlaLeuGlyMetGluSerGlyGluIleHisSerAspGlnIleThrAlaSerSer----- 295
QY 805 GAGTGGACTGACACACAGGGCAAGACAGTGTGGAACCAACCAAGCCAGGCTGAAA 864
Db 296 -----GlnYrGlyThrAsnTyrSerValGluArgSerArgLeuAsn 309
QY 865 AAACCTGGACCGCTTGGCTGCTTTCGCTGATGATGATGATGATGATGATGATGATG 924
Db 310 TyrProGluAsnGlyTyrThrProGlyGluAspSerYrLysGluTyrIleGlnValAsp 329
QY 925 TTGAATAAGAAAGAAATAACAGCATATATAACCACTGGA-----TCCACCATGCTG 978
Db 330 LeuGlyLeuLeuArgPheValThrAlaValGlyThrGlnGlyAlaIleSerLysGluThr 349
QY 979 GAGCAATATTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1038
Db 350 LysLysLysYrTyTyValLysThrYrArgValAspIleSerSerAsnGlyGluAspTyr 369
QY 1039 ACTGTGTACAGAGACCTGGTGTGAGCAGCAGATAGATATTTCAAGGAACAAGATTAT 1098
Db 370 IleSerLeuLysGluGly-----AsnLysAlaIlePheGlnGlyAsnThrAsnPro 387
QY 1099 CACCAAGGATGTCGTAATAACTTTTCCACCAATATTGACGCTTTTATTAGAGTGAAT 1158
Db 388 ThrAspValValLeuGlyValPheSerLysProLeuIleThrArgPheValArgIleLys 407
QY 1159 CCTACCCCAATGGCAGCAGAAATGGCCATGAAATGGAGCTGCTGCTGCTGCTGCTGCT 1218
Db 408 ProValSerTyrGluThrGlyIleSerMetArgPheGluValTyrGlyCys----- 424
QY 1219 CCTAAGTCTGCTCTCAAACTACTCAACCTCCA 1254
Db 425 -----LysIleThrAspTyrPro 430
RESULT 6
NR1_RAT
ID NR1_RAT
AC Q9QWJ9;
DT 16-Oct-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurotrophin-1 precursor (Vascular endothelial cell growth factor 165 receptor).
GN NR1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=97433085; PubMed=9288754;
RA Kolodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J., Ginty D.D.;
RT "Neurotrophin is a semaphorin III receptor";
RL Cell 90:753-762(1997).
CC -!- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF CERTAIN NEURONAL CIRCUITS AND IN ORGANOPHYSIS OUTSIDE THE NERVOUS SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT BINDS TO SEMAPHORIN 3A. THE PLGF-2 ISOFORM OF PGF, THE VEGF-165 ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS. IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: FOUND IN THE EMBRYONIC NERVOUS SYSTEM.
CC -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: AF016296; AAC53337.1; --
DR HSPP; P12259; ICZT.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR001092; HLH_Basic.
DR InterPro: IPR000998; MAM_domain.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR PRINTS: PR00020; MAMDOMAIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00231; FA58C; 2.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS50060; MAM_2; 1.
KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor.
FT SIGNAL 1 21
FT CHAIN 22 922
FT DOMAIN 22 855
FT TRANSMEM 856 880
FT DOMAIN 881 922
FT DOMAIN 27 141
FT DOMAIN 147 265
FT DOMAIN 275 424
FT DOMAIN 431 583
FT DOMAIN 645 811
FT DISULFID 27 54
FT DISULFID 82 104
FT DISULFID 147 173
FT DISULFID 206 228

QY 1180 ATTGCCATGAATGGAGCTGCTGGATGTCAGTTTATTCTAAAGTCTCTCCAAA 1239
 DDB 413 ValSerLeuArgPheGluValTyrGlyCys-----Lys 423
 QY 1240 CTACTCAACTCCA-----CCTCCTCGG 1263
 DDB 424 IleThrAspTyrProCysSerGlyMetLeuGlyMetValSerGlyLeuIleProAsp--- 442
 QY 1264 AACAGCAATGACCTCAAAACACATACAGCCCTCCAAATAGCAAGGTCGTGCCCA 1323
 DDB 443 -----SerGlnIleThrAlaSerThrGlnValAspArgAsnTrpIlePro 457
 QY 1324 AAATTTAGCAACCATACACCTCGAGTAGCAATGCTCCTGACAGACAGACAA 1383
 DDB 458 GluAsnAlaArgLeuIleThrSerArgSerGlyTyrAlaLeuProThrHisPro 477
 QY 1384 ACAACT 1389
 DDB 478 TyrThr 479

RESULT 8
 NRPL_XENLA STANDARD; PRT; 928 AA.
 ID AC P28824;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neupilin-1 precursor (A5 protein) (A5 antigen).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=9137458; PubMed=1908252;
 RA Takagi S., Hirata T., Agata K., Mochii M., Eguchi G., Fujisawa H.;
 RT "The A5 antigen, a candidate for the neuronal recognition molecule,
 RT has homologies to complement components and coagulation factors.";
 RL Neuron 7:298-307(1991).
 CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
 CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
 CC SIMILARITY). PRESUMED TO BE INVOLVED IN THE NEURONAL RECOGNITION
 CC BETWEEN THE OPTIC NERVE FIBERS AND THE VISUAL CENTERS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: RETINAL GANGLION CELLS AND VISUAL CENTER
 CC NEURONS.
 CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcements>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10467; BAA01260.1;
 CC HSP; P12259; ICDT
 CC InterPro; IPR000859; CUB_domain.
 CC InterPro; IPR000421; FA58_C.
 CC InterPro; IPR000998; MAM_domain.
 CC Pfam; PF00431; CUB; 2.
 CC Pfam; PF00629; MAM; 1.
 CC Pfam; PF00754; F5_F8_type_C; 2.
 CC PRINTS; PR00020; MAMDOMAIN.

DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; SM00137; MAM; 1.
 DR PROSITE; PS00740; MAM_1; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00060; MAM_2; 1.
 KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
 KW Antigen.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 928 NEUROPILIN-1.
 FT DOMAIN 22 860 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 861 883 POTENTIAL.
 FT DOMAIN 884 928 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 27 141 CUB 1.
 FT DOMAIN 147 265 CUB 2.
 FT DOMAIN 275 424 F5/8 TYPE C 1.
 FT DOMAIN 431 584 F5/8 TYPE C 2.
 FT DOMAIN 646 812 MAM.
 FT DISULFID 27 54 PROBABLE.
 FT DISULFID 82 104 PROBABLE.
 FT DISULFID 147 173 PROBABLE.
 FT DISULFID 206 228 PROBABLE.
 FT DISULFID 275 424 BY SIMILARITY.
 FT DISULFID 431 584 BY SIMILARITY.
 FT CARBOHYD 150 150 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 261 261 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 523 523 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 928 AA; 103416 MW; AF6B323B0A4C789D CRC64;

Alignment Scores:
 Pred. No.: 3,8e-20 Length: 928
 Score: 383.00 Matches: 138
 Percent Similarity: 38.88% Conservative: 70
 Best Local Similarity: 25.79% Mismatches: 197
 Query Match: 9.65% Indels: 130
 DB: 1 Gaps: 22

US-10-060-830-2 (1-2190) x NRPL_XENLA (1-928)

QY 13 CTCCTGCTTACTGTGTC-----CTGCTCTGCTGCTGAGGAGCGCTGGA 57
 DDB 1 MetLeuLeuArgLeuLeuSerCysCysStrpLeuLeuSerLeuArgSerStrp 20
 QY 58 GCCCAGCAAGTGTGGATGTGGACACACTGTACTAGCCCTGAGAGTGGAACCTTACA 117
 DDB 21 AlaSerArgAsnAspLysCysGlyAspThrIleLysIleThrSerProSerTyrLeuThr 40
 QY 118 TCCTATAAATACCCACAGACCTATCCCAACACAGACACTGTTTGTGAATGGAGATCCGTGA 177
 DDB 41 SerAlaGlyTyrProHisSerTyrProSerGlnArgCysGluTrpLeuIleGlnAla 60
 QY 178 ---AGATGGGAGAGAGTCCGCATCAAAATTGGT---GACTTGCATTCAGATTCT 231
 DDB 61 ProGluHisTyrGlnArgIleMetIleAsnProHisPheAspLeuGluAspArg, 80
 QY 232 GATCTTGTCTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 291
 DDB 81 Glu---CysLysTyrAspTyrValGluValIleAspGlyAspAsnAlaAsnGlyGlnLeu 99
 QY 292 ATAGGCAATACTCTGGTCTGGGTTCGAAATGAACCAATTCATCA---ATTGAATCAAAAGGC 348
 DDB 100 LeuGlyLysTyrCysGly-----LysIleAlaProSerProLeuValSerThrGly 116
 QY 349 AATGAATCACAATGCTCTTCTGATGAGTGAATTCATGCTGCTGCGCGGATTTTGTGCC 408
 DDB 117 ProSerIlePheIleArgPheValSerAspTyrGluThrProGlyAlaGlyPheSerIle 136
 QY 409 TCATCTCTCTGTATAGATAAACACAGATCATTAATCTTCTTTGGACACTGCATCAATTTT 468
 DDB PRINTS: PR00020; MAMDOMAIN.

Db	137	ArgTyrGluValPheLys	----- 147
Qy	469	TTGGAACCTCAGTTCAGTAAGTACTGCCACGCTGGTGTCTCTCCTCTCTTCTGAGATA	528
Db	143	ThrGlyProGlu	-----CysSerArgAsnPhenThrSerSer 154
Qy	529	TCGTGGAACAAT	-----CCTCATGGATATAGAGATCC-----TCG 564
Db	155	AsnGlyValIleLysSerProLysTyrProGluLysTyrProAsnAlaLeuGluCysThr	174
Qy	565	CCATTGTGCATGCTGGTGGCATGCAGAGTAGTG	----- 600
Db	175	TyrIleIlePheAlaProLysMetGlnIleValLeuGluPheGluSerPheGluLeu	194
Qy	601	-----TCAAACACGTTGGCGGCCCAATCAGT	-----GTTGTAATTAAGT 639
Db	195	GluAlaAspSerAsnAlaProGlyGlyGlnThrCysArgTyrAspTrpLeuGlyIleTrp	214
Qy	640	AAAGGTATTCCTC	-----TATTATGAAGTCTCTTGGCTAAC 675
Db	215	AspGlyPheProGlyValGlyProHisIleGlyArgTyrCysGlyGlnAsnThrProGly	234
Qy	676	AAGCTCACATCTGCTGGGACACATATCTACAAAGTCTTTTACATTTAAACACAAAGTGA	735
Db	235	ArgValArgSerPheThrGlyLeuSerMetIlePheHisThrAspSerAlaIleAla	254
Qy	735	----- 735	----- 735
Db	255	LysGluGlyPhePheAlaAsnPheSerValValGlnSerAsnThrAspGluAspPheGln	274
Qy	736	TGTTATGGAACTGGGGATGGAGTCTGGTGTGATCGCGGATCCTCAATAAACACAGCATTA	795
Db	275	CysLysGluAlaLeuGlyMetGluSerGlyGluIleHisPheAspGlnIleSerValSer	294
Qy	796	TCT-----GTGCTGGAGTGGACTGACCACACAGGGCAA	-----GAGAACAGT 837
Db	295	SerGlnTyrSerMetAsnTrpSerAlaGluArgSerArgLeuAsnTyrValGluAsnGly	314
Qy	838	TGGAACCCAAAAAGCCAGGCTGAAAAAACCTGGACCGCCTTGGGCTGCTTTGGCCACT	897
Db	315	TrpThrProGlyGluAspThrValLys	----- 323
Qy	898	GATGAATACACAGTGTACAAATAGATTGTGAATAAGGAAAAAGAAATAACAGCATTATA	957
Db	324	-----GluTrpIleGlnValAspLeuGluAsnLeuArgPheValSerGlyIleGly	340
Qy	958	ACCACATGGA-----TCCACCATGGTGGAGCACATTTACTATGTGCTGCTGCCTACAGATC	1011
Db	341	ThrGlnGlyAlaIleSerLysGluThrLysLysLysTyrPheValLysSerTyrLysVal	360
Qy	1012	CTGTACAGTGCATGATGGGCAGAAATGGAGCTGTGTACAGAGAGCCTGGTGTGGACCAAGAT	1071
Db	361	AspIleSerSerAsnGlyGluAspTrpIleThrLeuLysAspGly-----AsnLysHis	378
Qy	1072	AGATATTTCAGGAACAAAGATTTATCACCAGGATGTGCGTAATAACTTTTCCGCCACA	1131
Db	379	LeuValPheThrGlyAsnThrAspAlaThrAspValValTyrArgProPheSerLysPro	398
Qy	1132	ATTATTGCACGTTTATTAGATGGAATCCTACCCAATGGCAGCAGAAAATTGCCATGAAA	1191
Db	399	ValIleThrArgPheValArgLeuArgProValThrTrpGluAsnGlyIleSerLeuArg	418
Qy	1192	ATGGAGTGTCTCGGATGTCAAGTTTATCTCTAAAGTGTCTCTCCAAACACTTACTCAACT	1251
Db	419	PheGluLeuTyrGlyCys-----LysIleThrAspTyr	429
Qy	1252	CCACCTCCTCGGAACAGCAATGACCTC-----AAAAACACTACAGCC	1293
Db	430	ProCysSerArgMetLeuGlyMetValSerGlyLeuIleSerAspSerGlnIleThrAla	449
Qy	1294	CTCCCAAAATAGCCAAAGTGTGCGCCCAAAATTTACGCAACCACTACAACTCGCGAGT	1353
Db	450	SerSerGlnValAspArgAsnTrpValProGluLeuAlaArgLeuValThrSerArgSer	469

```

Qy 1354 -----AGCAATGAATTCTGCACACAGAA 1380
Db 470 GlyTrrPalaLeuProSerAsnThrHisProTyrThrLysGlu 484
      |||||
      |||||
      |||||

RESULT 9
FA5_PIG STANDARD; PRT; 2258 AA.
ID FA5_PIG AC Q9GLPI;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C
RP DOMAINS.
RC TISSUE=Liver;
RX MEDLINE=21121490; PubMed=11229814;
RA Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J.; Neame P.J.;
RA Kim H.K.W.;
RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional
RT protein modeling of membrane binding sites and comparative anatomy of
RT domains";
RL Cell. Mol. Life Sci. 58:148-159(2001).
CC -1- FUNCTION: Coagulation factor V is a cofactor that participates
CC with factor Xa to activate prothrombin to thrombin.
CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC chain, noncovalently bound. The interaction between the two chains
CC is calcium-dependent.
CC -1- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1
CC and C2 may be involved in membrane binding.
CC -1- PMF: Thrombin activates factor V proteolytically to the active
CC cofactor, factor Va (formation of a heavy chain at the N-
CC terminus and a light chain at the C-terminus).
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF191308; AAC28381.1; -.
DR HSSP; P12259; ICZT.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_Type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 2258 COAGULATION FACTOR V.
FT CHAIN 23 737 HEAVY CHAIN (BY SIMILARITY).
FT PEPTIDE 738 1611 ACTIVATION PEPTIDE (CONNECTING REGION)
(BY SIMILARITY).
FT CHAIN 1612 2258 LIGHT CHAIN (BY SIMILARITY).
FT DOMAIN 30 329 F5/8 TYPE A 1.
FT DOMAIN 30 193 PLASTOCYANIN-LIKE 1.
FT DOMAIN 203 329 PLASTOCYANIN-LIKE 2.

```


QY	1927	G---ATTGGCGAATGTGATGTTTAAAGAAATCCTTT	1960
Db	841	GGATCGGGAATGTGAATGTTTAAAGAAATCCTT	877
RESULT 2			
AL1378788/c			
LOCUS	AL1378788	712 bp	mRNA
DEFINITION	tc67d04.x1 Soares_NhMHPu.S1 Homo sapiens cDNA clone IMAGE:2069647		linear EST 18-MAR-1999
	3' similar to TR:Q14089 Q14089 HYPOTHETICAL PROTEIN ;, mRNA		
ACCESSION	AL1378788.1	GI:4188641	
VERSION	AL1378788		
KEYWORDS	EST.		
SOURCE	human,		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 712)		
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgaps-r@mail.nih.gov		
	This clone is available royalty-free through LLNL ; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Insert length: 861 Std Error: 0.00		
	Seq primer: -40UP from Gibco		
	High quality sequence stop: 461.		
FEATURES	Location/Qualifiers		
source	1. 712		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2069647"		
	/clone_lib="Soares_NhMHPu.S1"		
	/tissue_type="Pooled human melanocyte, fetal heart, and		
	pregnant uterus"		
	/lab_host="DH10B"		
	/note="Organ: mixed (see below); Vector: pT7T3D-Pac		
	(Pharmacia) with a modified polylinker; Site 1: Not 1;		
	Site 2: Eco RI; Equal amounts of plasmid DNA from three		
	normalized libraries (melanocyte 2NBHM, pregnant uterus		
	NhMHPu, and fetal heart NBH19w) were mixed, and ss cDNA		
	were made in vitro. Following HAP purification, this DNA		
	was used as tracer in a subtractive hybridization		
	reaction. The driver was PCR-amplified cDNAs from pools of		
	5,000 clones made from the same 3 libraries. The pools		
	consisted of 1.M.A.G.E. clones 260232-265223,		
	340488-345479, and 484488-489479."		
BASE COUNT	166 a 143 c 173 g 230 t		
ORIGIN			
Query Match	34.9%;	Score 694.8;	DB 9; Length 712;
Best Local Similarity	97.6%;	Pred. No. 6.5e-18;	
Matches 695;	Conservative 0;	Mismatches 17;	Indels 0; Gaps 0;
QY	586	GACCACAGGCGCAAGACAGACAGTTGGAAACCCCAAAAAGCCAGGCTGAAAAAACCTTGA	645
Db	712	GACCACCAAGCCAAAGAGACGACCTGGAAACACCAAAAGCCAGGCTGAAAAAACCTGAC	653
QY	646	CGCCCTTGGGCTGCTTTTGCCACTGATGAATACCAATACCAATAGATTGTAATGAAG	705
Db	652	CGCCCTTGGGCTGCTTTTGCCACTGATGAATACCAATAGATTGTAATGAAG	593
QY	706	GAAGAAGAAATACAGGCAATTATACCACTGGATCCACCATGGTGAGCCACCAATTACTAT	765
Db	592	GAAGAAGAAATACAGGCAATTATACCACTGGATCCACCATGGTGAGCCACCAATTACTAT	533
QY	766	GTGTCGTCCCTACAGATCCCTGTACAGTGTATGGSCAAATGGACTCTGTACAGAGAG	825
Db	532	GTGTCGTCCCTACAGATCCCTGTACAGTGTATGGSCAAATGGACTCTGTACAGAGAG	473

```

QY 826 CCTGGTGTGGAGCAAGATATTTTCAGGAAACAAGATTTATCACCAGGATGCGCT 885
Db 472 CTTGGTGTGGAGCAAGATATTTTCAGGAAACAAGATTTATCACCAGGATGCGCT 413
QY 886 AATAACTTTTGGCCACCAATATTGACAGCTTTTATTAGAGTGAATCCTACCCCAATGGCAG 945
Db 412 AATAACTTTTGGCCACCAATATTGACAGCTTTTATTAGAGTGAATCCTACCCCAATGGCAG 353
QY 946 CAGAAAATGGCCATGAAATGGAGCTGCTGGATGTCAGTTTATTCTTAAAGTCTGCTCT 1005
Db 352 CAGAAAATGGCCATGAAATGGAGCTGCTGGATGTCAGTTTATTCTTAAAGTCTGCTCT 293
QY 1006 CCAAACTTACTCAACTCCACCTCTCGGACAGCAATGACCTCAAAAACACTACAGCC 1065
Db 292 CCAAACTTACTCAACTCCACCTCTCGGACAGCAATGACCTCAAAAACACTACAGCC 233
QY 1066 CTCCAAAATACCCAAAGTGTGCCCCAAATTTACGCAACCACTACAACTCTGCAGT 1125
Db 232 CTCCAAAATACCCAAAGTGTGCCCCAAATTTACGCAACCACTACAACTCTGCAGT 173
QY 1126 AGCAATGAATTTCTGCACAGACAGCAACAACTGCCAGTCTCTGATATCAGAAATACT 1185
Db 172 AGCAATGAATTTCTGCACAGACAGCAACAACTGCCAGTCTCTGATATCAGAAATACT 113
QY 1186 ACCGTAACTCCAAATGTAAACCAAGATGTAGCGCTGGCTGCAGTTCTTGTCCCTGTGCTG 1245
Db 112 ACCGTAACTCCAAATGTAAACCAAGATGTAGCGCTGGCTGCAGTTCTTGTCCCTGTGCTG 53
QY 1246 GTCATGTCTCACTACTCTCATCTCTCATATTTAGTGTGCTGGCACTGGA 1297
Db 52 GTCATGTCTCACTACTCTCATCTCTCATATTTAGTGTGCTGGCACTGGA 1

RESULT 3
LOCUS BI860608 694 bp mRNA linear EST 10-OCT-2001
DEFINITION 603386804F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395609 5',
mRNA sequence.
ACCESSION BI860608
VERSION BI860608.1 GI:16001355
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapps@remail.nih.gov
Tissue Procurement: DCTP/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12008 row: g column: 02
High quality sequence stop: 692.
FEATURES
Location/Qualifiers
1..694
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5395609"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-Sport6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

```

```

BASE COUNT 230 a 165 c 147 g 152 t
ORIGIN
Query Match 34.8%; Score 683.4; DB 13; Length 694;
Best Local Similarity 99.1%; Pred. No. 1.6e-182;
Matches 687; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 514 GGAACACTGGGAGTGGAGTCTGTGTGATCGCGATCCTCAAAATACAGCATCATCTGTG 573
Db 2 GGACGGGTGGTGTGGAGTCTGTGTGATCGCGATCCTCAAAATACAGCATCATCTGTG 61
QY 574 CTGAGTGTGACTCACCACAGGCGAAGAACAGTTGGAAACCCCAAAAGCCAGCGTG 633
Db 62 CTGAGTGTGACTCACCACAGGCGAAGAACAGTTGGAAACCCCAAAAGCCAGCGTG 121
QY 634 AAAAAACCTGGAGCGCTTGGCTGCTTTTGGCCACTGATGAATACCAAGTGGTTACAATA 693
Db 122 AAAAAACCTGGAGCGCTTGGCTGCTTTTGGCCACTGATGAATACCAAGTGGTTACAATA 181
QY 694 GATTGAATGAAGAAAGAAATAACAGGCAATTAACCACTGGATCCACATGGTGGAG 753
Db 182 GATTGAATGAAGAAAGAAATAACAGGCAATTAACCACTGGATCCACATGGTGGAG 241
QY 754 CACAATTAATCTGCTGCTACAGATCCTGTACAGTGTATGATGCGAGAAATGACT 813
Db 242 CACAATTAATCTGCTGCTACAGATCCTGTACAGTGTATGATGCGAGAAATGACT 301
QY 814 GTGTACAGAGAGCTGTGTGGAGCAAGATAAGATATTCAAGGAAACAAGATTTATCAC 873
Db 302 GTGTACAGAGAGCTGTGTGGAGCAAGATAAGATATTCAAGGAAACAAGATTTATCAC 361
QY 874 CAGATGTGCGTAATAACTTTTGGCCACCAATATTTCAGCTTTTATTAGAGTAATCCT 933
Db 362 CAGATGTGCGTAATAACTTTTGGCCACCAATATTTCAGCTTTTATTAGAGTAATCCT 421
QY 934 ACCAATGGCAGAGAAATTTGCCATGAATGGAGCTGCTCGGATCTCAGTTTATTCCT 993
Db 422 ACCAATGGCAGAGAAATTTGCCATGAATGGAGCTGCTCGGATCTCAGTTTATTCCT 481
QY 994 AAAGTGTCTCTCCAAAACCTTACTCAACCTCCACCTCTCGGAACAGCAATGACCTCAA 1053
Db 482 AAAGTGTCTCTCCAAAACCTTACTCAACCTCCACCTCTCGGAACAGCAATGACCTCAA 541
QY 1054 AACACTACAGCCCTCCAAAATAGCCAAAGTGTGTCGCCCAAAATTTAGCAACACTA 1113
Db 542 AACACTACAGCCCTCCAAAATAGCCAAAGTGTGTCGCCCAAAATTTAGCAACACTA 601
QY 1114 CAACCTCGCAGTAGCAATGAATTTCTGCAACAGACAGCAACAACTGCCAGTCTGAT 1173
Db 602 CAACCTCGCAGTAGCAATGAATTTCTGCAACAGACAGCAACAACTGCCAGTCTGAT 661
QY 1174 ATCAGAAATTAACCTGTAACCTCAAAATGTAACC 1206
Db 662 ATCAGAAATTAACCTGTAACCTCAAAATGTAACC 694

RESULT 4
LOCUS AI435602/c
DEFINITION th79c10.x1 Soares.NHMPu.S1 Homo sapiens cDNA clone IMAGE:2124882
3' similar to TR:Q14089 Q14089 HYPOTHETICAL PROTEIN ;, mRNA
sequence.
ACCESSION AI435602
VERSION AI435602.1 GI:4304334
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

```

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs@rcmail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 865 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 450.

FEATURES
source

1. .641
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2124882"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3-Pac
(Pharmacia) with a modified polylinker; Site: 1: Not I;
Site: 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NHM, pregnant uterus
NHMPu, and fetal heart NHHL19) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT
ORIGIN

Query Match 31.8%; Score 623.2; DB 9; Length 641;
Best Local Similarity 99.2%; Pred. No. 1.9e-165;
Matches 637; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 556 CTGCTTTTCCACTGATGATACAGTGGTTACAAATAGATTTGAATGAAGAAAGAAA 715
DB 640 CTGCTTTTCCCTCTGATGATACAGTGGTTACAAATAGATTTGAATGAAGAAAGAAA 581
QY 716 TAACAGGATATATACCACTGGATCCACCATGGTGGAGCAAAATACATGCTGCTGCT 775
DB 580 TACAGGATATATACCACTGGAT-ACCATGGTGGAGCAAAATACATGCTGCTGCT 523
QY 776 ACAGATCTCTACAGTATGATGGCAGAAATGGATGCTGTACAGAGAGCGCTGTGG 835
DB 522 ACAGATCTCTACAGTATGATGGCAGAAATGGATGCTGTACAGAGAGCGCTGTGG 463
QY 836 AGCAAGATAAGATATTTCAAGGAACAAAGATATATCACCAGATGTGGTAACTATTT 895
DB 462 AGCAAGATAAGATATTTCAAGGAACAAAGATATATCACCAGATGTGGTAACTATTT 403
QY 896 TGCCACCAATATTTCAGCTTTTATTAGAGTGAATCCCTACCAATGGCAGCAAAATTTG 955
DB 402 TGCCACCAATATTTCAGCTTTTATTAGAGTGAATCCCTACCAATGGCAGCAAAATTTG 343
QY 956 CCATGAAATGGAGTGTGCGATGCTAGTTTATTCCTTAAAGTGTGCTCTCCAAACATTA 1015
DB 342 CCATGAAATGGAGTGTGCGATGCTAGTTTATTCCTTAAAGTGTGCTCTCCAAACATTA 283
QY 1016 CTCACCTCCACCTCTCGGACACGAATGACCTCAAAACACTACAGCCCTCCAAAAA 1075
DB 282 CTCACCTCCACCTCTCGGACACGAATGACCTCAAAACACTACAGCCCTCCAAAAA 223
QY 1076 TAGCCAAAGGTGTGCCCCCAAAATTTAGCCACCACTACAACTCGCAGTAGCAATGAAT 1135
DB 222 TAGCCAAAGGTGTGCCCCCAAAATTTAGCCAACTACAACTCGCAGTAGCAATGAAT 163
QY 1136 TTCTTCGACAGACAGAAACAACTGCGAGTCTGTATATCAGAAATACCTACCGTAACATC 1195
DB 162 TTCTTCGACAGACAGAAACAACTGCGAGTCTGTATATCAGAAATACCTACCGTAACATC 103
QY 1196 CAAATGTACCAAGATGTAGCGCTGGCTGCGAGTTCTTCTCCTGTGGTCAATGGTCC 1255

Db 102 CAAATGTACCAAGATGTAGCGCTGGCTGCAGTCTTCTGTCCCTGTGCTCATGGTCC 43
QY 1256 TCACACTCTCATCTTCATCATATTAGTGTGCTGGCACTGGA 1297
DB 42 TCACACTCTCATCTTCATCATATTAGTGTGCTGGCACTGGA 1

RESULT 5

BQ014739/c
LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. .680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5833547"
/clone_lib="NCI_CGAP_Ed1"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Left Pubic Bone; Vector: pT7T3-Pac
(Pharmacia) with a modified polylinker; Site: 1: EcoR I;
Site: 2: Not I; NCI_CGAP_Ed1 is a normalized cDNA library
containing the following tissue(s): Chondrosarcoma cell
line C95. The library was constructed according to Bonaldo
, Lennon and Soares, Genome Research, 6:791-806, 1996.
First strand cDNA synthesis was primed with an oligo-dT
primer containing a Not I site. Double stranded cDNA was
ligated to an EcoR I adaptor, digested with Not I, and
cloned directionally into pT7T3-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is GCTCAGGCT.
TAG_LIB=UI-H-Ed1
TAG_TISSUE=chondrosarcoma
TAG_SEQ=CGTCAAGCT"

BASE COUNT
ORIGIN

Query Match 31.3%; Score 613.4; DB 14; Length 680;
Best Local Similarity 99.8%; Pred. No. 1.2e-162;
Matches 614; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1348 CGGGCAGGTTGGTGAAGGAATGAAGCAGTTCTTCTGCAAGCACTGGACCATGAG 1407
DB 680 CGGGCAGGTTGGTGAAGGAATGAAGCAGTTCTTCTGCAAGCACTGGACCATGAG 621
QY 1408 GAAACCCAGTTCGCTATAGCAGCGAAGTTAATCACCTGAGTCAAGAGAAGTCAAC 1467

```
Db 620 GAAACCCAGTTCGCTATAGCAGCAGCGAAGTTAATCACTGAGTCCAGAGAGTCACC 561
QY 1468 ACAGTGTGCTGAGCTGACCTCTGAGAGTAGTGTCTGACCCACTGTGTAGAGAAATGTGGT 1527
Db 560 ACAGTGTGCTGAGCTGACCTCTGAGAGTAGTGTCTGACCCACTGTGTAGAGAAATGTGGT 501
QY 1528 ACATCTCATCAAGATCTACCTTTAAACAGAGAGAGAGAAAGAACAGCTATGAGAC 1587
Db 500 ACATCTCATCAAGATCTACCTTTAAACAGAGAGAGAGAAAGAACAGCTATGAGAC 441
QY 1588 CTAGATCTCTTACAACTACACAGGCGCAGGAGTGTATCATGCTCTATGCTGAACCACTCCCA 1647
Db 440 CTAGATCTCTTACAACTACACAGGCGCAGGAGTGTATCATGCTCTATGCTGAACCACTCCCA 381
QY 1648 ATTACGGGGCTGAGTATGCAACCCCAATCATCATGACATGTCAGGGCAGCCCACT 1707
Db 380 ATTACGGGGCTGAGTATGCAACCCCAATCATCATGACATGTCAGGGCAGCCCACT 321
QY 1708 TCAGTGTGCTGAGCTCTCCACATCCACTTCAAGGCTACGGGGAACCACTCCCACT 1767
Db 320 TCAGTGTGCTGAGCTCTCCACATCCACTTCAAGGCTACGGGGAACCACTCCCACT 261
QY 1768 GTGGGAACCTTACAACTACACTTCTCCAGGACTGACAGCTGCTCCCTCAGCCAGGCCAG 1827
Db 260 GTGGGAACCTTACAACTACACTTCTCCAGGACTGACAGCTGCTCCCTCAGCCAGGCCAG 201
QY 1828 TATGATACCCGGAAGCTGGGAAGCAGGCTCTACCTGCCCCAGACAAATGGTGTACCAG 1887
Db 200 TATGATACCCGGAAGCTGGGAAGCAGGCTCTACCTGCCCCAGACAAATGGTGTACCAG 141
QY 1888 GTCCACAGAGCACAAAGAGTATCAGGAGCAGGAGGATGGGAATGTGATGTTTT 1947
Db 140 GTCCACAGAGCACAAAGAGTATCAGGAGCAGGAGGATGGGAATGTGATGTTTT 81
QY 1948 AAGAAATCCTTTGA 1962
Db 80 AAGAAATCCTTTGA 66
```

```
RESULT 6
AW383889 664 bp mRNA linear EST 04-FEB-2000
LOCUS QV3-HT0364-241199-024-b01 HT0364 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW383889
ACCESSION AW383889
VERSION AW383889.1 GI:6888548
KEYWORDS EST.
SOURCE human.
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 664)
HGCP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
```

```
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3&2=QV3-HT0364-
241199-024-b01&t3=1999-11-24&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 53
High quality sequence stop: 623.
Location/Qualifiers
1..664
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```
FEATURES
source
```

```
/clone.lib="HT0364"
/dev.stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 185 a 130 c 157 g 192 t
ORIGIN
```

```
Query Match 30.0%; Score 588.6; DB 10; Length 664;
Best Local Similarity 97.7%; Pred. No. 1.3e-155;
Matches 597; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 99 CCAATTCATTAATGAATCAAAAGCAATGAATCAATCAATGCTGTCATGAGTGAATCCATGT 158
Db 46 CCTTCGATTAATCAAAAGCAATGAATCAATCAATGCTGTCATGAGTGAATCCATGT 105
QY 159 TTCTGACGCGGATTTTGGCCCTCATCTCTGTATAGATAAACAAGATCTAATTACTTG 218
Db 106 TTCTGACGCGGATTTTGGCCCTCATCTCTGTATAGATAAACAAGATCTAATTACTTG 165
QY 219 TTGGAACACTGTCATCAAAATTTTGGAACTGAGTTTCAGTAAGTACTGCCAGCTGGTTG 278
Db 166 TTGGAACACTGTCATCAAAATTTTGGAACTGAGTTTCAGTAAGTACTGCCAGCTGGTTG 225
QY 279 TCTGCTTCCTTTTGGCTGAGATATCTGGAACAATTCCTCATGATATAGAGATTCCTGCC 338
Db 226 TCTGCTTCCTTTTGGCTGAGATATCTGGAACAATTCCTCATGATATAGAGATTCCTGCC 285
QY 339 ATTGTCATGCGCTGGTGTGATCGAGGAGTAGTGTCAACACAGTGGGGCGGCCAAATCAG 398
Db 286 ATTGTCATGCGCTGGTGTGATCGAGGAGTAGTGTCAACACAGTGGGGCGGCCAAATCAG 345
QY 399 TGTGTAATTAAGTAAAGGATATCCCTATTATGAAGTCTTTTGGCTTAAACAACGTCACATC 458
Db 346 TGTGTAATTAAGTAAAGGATATCCCTATTATGAAGTCTTTTGGCTTAAACAACGTCACATC 405
QY 459 TGTGTTGGGACACTTATCTACAAGTCTTTTACATTTTAAACAAGTGGATGTATAGGAAC 518
Db 406 TGTGTTGGGACACTTATCTACAAGTCTTTTACATTTTAAACAAGTGGATGTATAGGAAC 465
QY 519 ACTGGGATGGAGTGTGCTGATCGCGGATCCTCAATAACACCATCTCTGCTGGA 578
Db 466 ACTGGGATGGAGTGTGCTGATCGCGGATCCTCAATAACACCATCTCTGCTGGA 525
QY 579 GTGACTGTACACACAGGCGCAAGAGACAGTGGAAACCCCAAAAGCCAGGCTGAAAAA 638
Db 526 GTGACTGTACACACAGGCGCAAGAGACAGTGGAAACCCCAAAAGCCAGGCTGAAAAA 585
QY 639 ACCTGGACCGCTTGGGCTGCTTTTGGCACTGATGATACCAAGTGGTTACAAATAGATT 698
Db 586 ACCTGGACCGCTTGGGCTGCTTTTGGCACTGATGATACCAAGTGGTTACCAATAGATT 645
QY 699 GAATAAGGAAA 709
Db 646 GATTACGGAAA 656
```

```
RESULT 7
AA147037/c
LOCUS AA147037
DEFINITION z032a05.s1 Stragatene colon (#937204) Homo sapiens cDNA clone
IMAGE:588560 3', similar to TR:G704441 G704441 HYPOTHETICAL 40.0 KD
PROTEIN ;, mRNA sequence.
ACCESSION AA147037
VERSION AA147037.1 GI:1716444
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 671)
Hillier L., Lennon G., Becker M., Bonaldo M.F., Chiapelli B.,
Chisoe S., Dietrich N., Dubuque T., Favello A., Gish W., Hawkins
M., Hultman M., Kucaba T., Lacy M., Le M., Le N., Mardis E., Moore
B., Morris M., Parsons J., Prange C., Rifkin L., Rohlfing T.,
Schellenberg K., Soares M.B., Tan F., Thierry-Mieg J., Trevaskis E.,
Underwood K., Wohldmann P., Waterston R., Wilson R. and Marra M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 439.
FEATURES
source
1..671
/organism="Homo sapiens"
/db_xref="GDB:4620862"
/db_xref="taxon:9606"
/clone="IMAGE:58580"
/clone_lib="Stratagene colon (#937204)"
/tissue_type="tumor"
/cell_line="T84 carcinoma cell line"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: colon; Vector: pBluescript SK-; Site:1:
EcORI; Site:2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5'
CTCGAGCTTTTCTTTTCTTTT 3' 2 others
BASE COUNT 176 a 162 c 133 g 198 t
ORIGIN
Query Match 29.8%; Score 584.8; DB 9; Length 671;
Best Local Similarity 97.9%; Pred. No. 1.6e-154;
Matches 645; Conservative 0; Mismatches 8; Indels 6; Gaps 5;
QY 287 CTTTGTGATGATCTGGAACAAATTCCTCATGGATATAGAGATTCCTCGCATTTGCA 346
DB 653 CTTTGTGATGATCTGGAACAAATTCCTCATGGATATAGAGATTCCTCGCATTTGCA 598
QY 347 TGGCTGTGCTGATGAGGAGTAGTGTCAACACAGTTTGGGGGCGGCAATCAGTGTGTAA 406
DB 597 TGGCTGTGCTGATGAGGAGTAGTGTCAACACAGTTTGGGGGCGGCAATCAGTGTGTAA 540
QY 407 TTAGTAAGGTATTCCTTATATGAAGTCTTTTGGGTACACAGTGCACATCTGTGTGG 466
DB 539 TTAGTAAGGTATTCCTTATATGAAGTCTTTTGGGTACACAGTGCACATCTGTGTGG 480
QY 467 GACACTTATCTACAAGTCTTTTACATTTAAGACAAGTGTGTTAGGACACTGGGA 526
DB 479 GACACTTATCTACAAGTCTTTTACATTTAAGACAAGTGTGTTAGGACACTGGGA 420
QY 527 TGGAGTCTGTGTGATCGCGGATCCTCAATAACAGCATCATCTGTGCTGGAGTGGACTG 586
DB 419 TGGAGTCTGTGTGATCGCGGATCCTCAATAACAGCATCATCTGTGCTGGAGTGGACTG 360
QY 587 ACCACAGGGCAGAGAACAGTTGGAAACCCCAAAAGCCAGGCTGAAACACCTGGAC 646
DB 359 ACCACAGGGCAGAGAACAGTTGGAAACCCCAAAAGCCAGGCTGAAACACCTGGAC 300
QY 647 CGCCTTGGGTGCTTTGGCCACTGATGAATACAGTGGTTACAAATAGATTTGAATAGG 706
DB 299 CCCCTTGGGTGCTTTTGGCCACTGATGAATACAGTGGTTACAAATAGATTTGAATAGG 240
QY 707 AAAAGAAAATACAGGCATTATACCACTGGATCCACCATTGGTGGAGCAACAATTTACTATG 766

Db 239 AAAGAAAATACAGGCATTATACCACTGGATCCACCATTTACTATG 180
QY 767 TGCTGCTCTACAGAAATCCTGTACAGTGTGATGGGAGAAATGAGTGTGTACAGAGC 826
Db 179 TGCTGCTCTACAGAAATCCTGTACAGTGTGATGGGAGAAATGAGTGTGTACAGAGC 120
QY 827 CTGTGTGGGAGCAGATTAAGATATTTCAAGGAAACAAAGATTTATCACAGGATCTGGTA 886
Db 119 CTGTGTGGGAGCAGATTAAGATATTTCAAGGAAACAAAGATTTATCACAGGATCTGGTA 60
QY 887 ATAACTTTTGGCCACCAATTTATTCAGCTTTTATTAGAGTGAATCTTACCAATGGCAG 945
Db 59 ATAACTTTTGGCCACCAATTTATTCAGCTTTTATTAGAGTGAATCTTACCAATGGCAG 1
RESULT 8
BE94226
LOCUS BE94226 618 bp mRNA linear EST 20-OCT-2000
DEFINITION 601437167F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922169 5',
mRNA sequence.
ACCESSION BE94226
VERSION BE94226.1 GI:10356380
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 618)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Prepared by: The I.M.A.G.E. Consortium, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9756 row: e column: 18
High quality sequence stop: 618.
FEATURES
source
1..618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3922169"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Skin; Vector: pCMV-SPORT6; Site:1: Noti;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 201 a 135 c 139 g 143 t
ORIGIN
Query Match 29.8%; Score 584; DB 12; Length 618;
Best Local Similarity 99.7%; Pred. No. 2.6e-154;
Matches 606; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 479 CAAGTCTTTTACATTTAAGACAGTGGATGTTATGGAACACTGGGATGGAGTCTGGT 538
DB 1 CAAGTCTTTTACATTTAAGACAGTGGATGTTATGGAACACTGGGATGGAGTCTGGT 60
QY 539 TGATCGGGGATCCTCAATAACAGCATCATCTGTGCTGGAGTGGACTGACACACAGGGC 598
DB 61 TGATCGGGGATCCTCAATAACAGCATCATCTGTGCTGGAGTGGACTGACACACAGGGC 120
QY 599 AAGAGACAGTTGGAAACCCCAAAAGCCAGGCTGAAACACCTGGACCGCTTGGGCTG 658
DB 121 AAGAGACAGTTGGAAACCCCAAAAGCCAGGCTGAAACACCTGGACCGCTTGGGCTG 180


```

Db 118 AATACTACCGTAACCTCAAAATGTAACCAAGATGTAGCGCTGGCTGCAGTCTTCTGTCCCT 59
QY 1240 GTGCTGGTCATGGTCTCTACTACTCTCATCTCATTATAGTGTGCTGGCAGCTGGA 1297
Db 58 GTGCTGGTCATGGTCTCTACTACTCTCTCTCTCATATTAGTGTGCTGGCAGCTGGA 1

RESULT 14
AV603144/c 575 bp mRNA linear EST 27-NOV-2001
LOCUS AV603144 Bos taurus kidney fetus Bos taurus cDNA clone EIK1014D09
DEFINITION 5', mRNA sequence.
ACCESSION AV603144
VERSION AV603144.1 GI:9725470
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 575)
AUTHORS Takasuga, A., Hirotsune, S., Itoh, R., Jitohzono, A., Suzuki, H., Aso, H.,
and Sugimoto, Y.
TITLE Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE 21570554
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cococ.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
FEATURES
Location/Qualifiers
1..575
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="EIK1014D09"
/tissue_type="kidney"
/dev_stage="fetus"
/lab_host="DH10B"
/notes="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 131 a 137 c 129 g 178 t
ORIGIN
Query Match 25.2%; Score 494; DB 10; Length 575;
Best Local Similarity 92.0%; Pred. No. 8.8e-129;
Matches 521; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 520 CTGGGATGAGTCTGGTGTGATCGCGGATCTCAATAACAGCATCATCTGCTGGAG 579
Db 575 CTAGGCATGGAATCTGGGTGTGATCGCTGATGCTCAATAACAGCATCATCTGCTGGAG 516
QY 580 TGGACTGACCACACAGGCGAAGAGAACAGTTGGAAACCCAAAGCCAGGCTGAAAAA 639
Db 515 TGGACGGACCACAGAGACAGAGAACAGCTGGAAACCTGGAAGCCAGGCTGAAAAA 456
QY 640 CTTGGACCGCTTGGGCTGCTTTTGGCACTGATGAATACCAAGTGGTTACAAATAGATTG 699
Db 455 CTTGGACCTCTTGGGCTGCTTTTGGCACTGATGAATATCATGCTGGTTGCAAGTAGATCTG 396
QY 700 AATAAGAAAGAAATAACAGGCATTATAACCACTGGATCCACATGTTGGGAGACAAT 759
Db 395 AATAAGAAAGAAAGATACAGGCATTGTGACTACTGGCTCCACCATGGTGGGAGACAAT 336
QY 760 TACTATGTGCTCGCTACAGAAATCCCTGTACATGTGATGGCGCAATAGGACTGTGTAC 819
||||| ||||| ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 335 TACTACGCTGTCAGCCTACAAGATTCTATACAGTGTCTATGGCAGAGGTGGACTGTGTAC 276
QY 820 AGAGAGCCTGGTGTGGAGCAAGATAGATATTTCAAGGAACAAGAAATATCACCAGGAT 879
Db 275 AGAGAGCCTGGTGTGGAGCAGGATAGATATTTCAAGGAACAAGAAATATCACCAGGAT 216
QY 880 GTGCGTAATAACTTTTGGCCACCAATATTGCACGTTTTATTAGAGTGAATCTACCCAA 939
Db 215 GTGCGTAATAACTTTTGGCCACCAATATTGCACGTTTTATTAGAGTGAATCTACCCAA 156
QY 940 TGGCAGCAGAAAATGCCATGAAAATGGAGTGTGCGGATGTCAAGTTTATTCCTAAAGT 999
Db 155 TGGCAGCAGAAAATGCCATGAAAATGGAGTGTGCGGATGTCAAGTTTATTCCTAAAGT 96
QY 1000 CGTCTCTCAAAAATCTACTCAACTCCACTCCCTCGGAAACAGCAATCACTCAAAAACACT 1059
Db 95 CGTCTCTCAAAAATCTACTCAACTCCCTCGGAGCAATCACTCAAAAATCACT 36
QY 1060 ACAGCCCTCCCAAAAATAGCCAAAGG 1085
Db 35 ACACCCCTCCCAAAAATAGCCAAAGG 10

RESULT 15
BF074573/c 560 bp mRNA linear EST 25-APR-2001
LOCUS BF074573
DEFINITION 222024 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF074573
VERSION BF074573.1 GI:10868084
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 560)
AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fehrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,
Perteau, G., Holt, I., Karameycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith, TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR PRIMERS
FORWARD: AGCAACAGCTATGACCAT
BACKWARD: GTTTCCAGTCACGACG
Plate: 81 row: M column: 18
Seq primer: ATTTAGTGACACTAG.
FEATURES
Location/Qualifiers
1..560
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT 141 a 131 c 113 g 175 t
ORIGIN

```

Query Match 25.0%; Score 490.2; DB 12; Length 560;
Best Local Similarity 92.3%; Pred. No. 1e-127;
Matches 516; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 470 ACTTATCTACAAGTCCTTTTACATTTAAGACAAAGTGGATGTTATGGAACACTGGGATGG 529
|||||
Db 560 ACTTATCTACAAGTCCTTTTACATTTAAGACAAAGTGGTGTATGGAACACTAGGCATGG 501
|||||
QY 530 AGTCTGTGTGATCGCGGATCCCTCAATAACAGCATCATCTGTGCTGGAGTGGACTGACC 589
|||||
Db 500 AATCTGGCGTGTGATGATGCTCAATAACAGCATCACTGTGTGGAGTGGACGGACC 441
|||||
QY 590 ACACAGGGCAAGAGACAGTGTGAAACCCCAAAAGCCAGGCTGAAAAACCTGGACCCG 649
|||||
Db 440 ACACAGGACAAGAGACAGTGTGAAACCCCTGAAAAAGCCAGGCTGAAAAACCTGGACCTC 381
|||||
QY 650 CTTGGGCTGCTTTTGGCCACTGATACCACTGGTGTACAAATAGATTTGAATAAGGAAA 709
|||||
Db 380 CTTGGGCTGCTTTTGGCCACTGATGAATATCAGTGGTGGCAAGTAGATCTGAAATAAGGAAA 321
|||||
QY 710 AGAAATAACAGGCATTATAACCACTGGATCCACCATTGGTGGAGCACAAATTACTATGTGT 769
|||||
Db 320 AGAATACACAGCATTTGTGACTACTGGCTCCACCATTGGTGGAGCACAAATTAATACGTGT 261
|||||
QY 770 CTGCTACAGATCCCTGTACAGTGTATGGGAGAAATGGACTGTGTACAGAGAGCCCTG 829
|||||
Db 260 CAGCCTACAGATCTATACAGTGTGTATGGGAGAGGTGGACTGTGTACAGAGAGCCCTG 201
|||||
QY 830 GTCTGGAGCAAGATAAGATATTTCAAGGAACAAGATATATCACCAGGATGTGCGTAATA 889
|||||
Db 200 GTGTGGACCAAGGATAAGATATTTCAAGGAACAAGATATATCACCAGGATGTGCGTAATA 141
|||||
QY 890 ACTTTTGGCCACCAATTATTGACGTTTATTAGAGTGAATCCTACCCAATGGCAGCAGA 949
|||||
Db 140 ACTTTTGGCCACCAATTATTGACGTTTATTAGAGTGAATCCTACCCAATGGCAGCAGA 81
|||||
QY 950 AAATTGCCATGAAATGGAGCTGCTCGGATGTCAGTGTATTCCTAAAGGTGCTCCTCCAA 1009
|||||
Db 80 AAATTGCCATGAAATGGAGCTGCTCGGATGTCATATTTATTATTAAGGTGCTCCTCCAA 21
|||||
QY 1010 AACTTACTCAACCTCCACC 1028
|||||
Db 20 AACTTACTCAACCTCGGCC 2

Search completed: January 21, 2003, 09:21:27
Job time : 2768.06 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 09:21:35 ; Search time 33.5541 seconds
(without alignments)
11242.484 Million cell updates/sec

Title: US-10-060-830-1113
Perfect score: 3559
Sequence: 1 atgctctgttctctctgct.....tttttaagaagaatctttga 1962

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/crn2.1/USPTO_spool/US10060830/runat_16012003_091239_26859/app_query.fasta_1.6933
-DB=PR73 -QFWT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn.1.1.166.ernat.16012003_091239_26859 -NCPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PR73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	361	10.1	2211	1 KFB05	coagulation factor
2	355.5	10.0	2183	2 T42764	coagulation factor
3	351	9.9	2224	1 KFH05	coagulation factor
4	349	9.8	401	2 S65138	glycoprotein antlg
5	349	9.8	427	2 S74211	PAS-6/7 protein pr
6	343	9.6	218	2 A47285	milk fat globule p
7	340	9.6	409	2 T11743	pp47 protein - pig
8	334	9.4	427	2 JC4915	ags protein precu
9	322	9.0	2319	2 A47004	coagulation factor
10	321	9.0	463	1 A36479	milk fat globule m
11	316	8.9	2351	1 EZHU	coagulation factor
12	308.5	8.7	216	2 A44258	factor VIII-assoc1
13	307	8.6	2133	2 T42763	coagulation factor
14	258	7.2	927	1 JQ0948	A5 antigen precurs

15	224.5	6.3	845	2	JC5256	adipocyte transcri
16	215.5	6.1	3133	2	S52093	hemocytin - silkwo
17	207.5	5.8	1283	2	T13799	neurexin IV - frui
18	204	5.7	719	2	S51739	transcription repr
19	200.5	5.6	855	2	S42621	protein-tyrosine k
20	187	5.3	913	2	A48280	receptor tyrosine
21	183.5	5.2	910	2	A53137	tyrosine kinase re
22	171	4.8	819	2	I48859	tyro 10 receptor k
23	168	4.7	876	2	A49508	protein-tyrosine k
24	145.5	4.1	737	2	T15615	hypothetical prote
25	145	4.1	1381	2	T31083	paranodin - rat
26	145	4.1	1385	2	T14158	neurexin IV - mous
27	135.5	3.8	737	2	T31349	hypothetical prote
28	127	3.6	837	1	A29512	LDL receptor precu
29	126.5	3.6	854	1	QRHYLD	LDL receptor precu
30	126.5	3.6	1843	2	S18803	collagen alpha 1(V
31	124.5	3.5	996	2	JE0237	apolipoprotein E r
32	124.5	3.5	2715	2	T13049	eyelid - fruit fly
33	124.5	3.5	3570	2	A45025	coagulation factor
34	123.5	3.5	1019	2	A38738	probable tenascin
35	122.5	3.4	4006	2	T09070	mucin MUC5B, trach
36	121.5	3.4	678	2	T49984	bromodomain protei
37	121.5	3.4	1827	2	T34288	hypothetical prote
38	121.5	3.4	3020	2	A43932	mucin 2 precursor,
39	120.5	3.4	760	2	F86387	probable Pto kinas
40	120.5	3.4	862	1	QRMSLD	LDL receptor precu
41	119.5	3.4	699	2	I38073	nucleolar phosphop
42	119.5	3.4	1203	2	T17415	mycelial surface a
43	119	3.3	786	2	T02729	serine/threonine-s
44	119	3.3	1140	2	D88690	protein F41H10.3 l
45	119	3.3	1611	2	T38236	hypothetical prote

ALIGNMENTS

RESULT 1

KFB05

coagulation factor V precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
C:Accession: A42580; A36457
R:Guinto, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A:Title: The complete cDNA sequence of bovine coagulation factor V.
A:Reference number: A42580; MUID:92147638; PMID:1737753
A:Accession: A42580
A:Molecule type: mRNA
A:Residues: 1-2211 <GUI>
A:Cross-references: GB:M81440; NID:G163037; PID:AAA30512.1; PID:G163038
A:Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBIP:80776)
R:Kalafatis, M.; Jenny, R.J.; Mann, K.G.
J. Biol. Chem. 265, 21580-21589, 1990
A:Title: Identification and characterization of a phospholipid-binding site of bovine
A:Reference number: A36497; MUID:91072354; PMID:2254316
A:Accession: A36497
A:Molecule type: protein
A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584, 1673-1676, 'X', 1678-1679, 'X', 1681, 'X'
R:Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.
Biochemistry 33, 13109-13116, 1994
A:Title: Determination of the disulfide bridges in factor Va heavy chain.
A:Reference number: A55979; MUID:95034740; PMID:7947716
A:Contents: annotation
A:Note: 566-Cys and 617-Cys were shown to have free sulphydryls
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa p
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasm
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-2211/Product: coagulation factor V #status predicted <MAT>
F:29-741/Product: coagulation factor Va heavy chain #status predicted <VAH>
F:29-345/Domain: A1 <DAL>

C:Species: Homo sapiens (man)
 C:Date: 19-May-1989 #sequence_revision 02-Jun-1995 #text_change 08-Dec-2000
 C:Accession: A56172; A42344; A28028; A27498; A25897
 R:Cripe, L.D.; Moore, K.D.; Kane, W.H.
 Biochemistry 31, 3777-3785, 1992
 A:Title: Structure of the gene for human coagulation factor V.
 A:Reference number: A42344; MUID:92232668; PMID:1567832
 A:Accession: A56172
 A:Molecule type: DNA
 A:Residues: 1-2224 <CRI>
 A:Cross-references: GB:J05368
 A:Accession: A42344
 A:Molecule type: DNA
 A:Residues: 48-58;79-89;120-130;191-201;239-249;313-323;368-378;428-437;461-471;533-542;
 2070-2111-2120;2172-2181 <CR2>
 R:Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kaufm
 Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
 A:Title: Complete cDNA and derived amino acid sequence of human factor V.
 A:Reference number: A28028; MUID:87260886; PMID:3110773
 A:Accession: A28028
 A:Molecule type: mRNA
 A:Residues: 1-857; R', 859-864, R', 866-924, 'E', 926-1763, 'I', 1765-2212, 'T', 2214-2224 <JEN>
 A:Cross-references: GB:M16967
 A:Note: parts of this sequence, including the amino end of the mature protein, were dete
 R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.
 Biochemistry 26, 6508-6514, 1987
 A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum
 A:Reference number: A27498; MUID:88107560; PMID:2827731
 A:Accession: A27498
 A:Molecule type: mRNA
 A:Residues: 1-1284, 'I', 1286-1600 <KAN>
 A:Cross-references: GB:M17785
 A:Note: parts of this sequence were determined by protein sequencing
 R:Kane, W.H.; Davie, E.W.
 Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
 A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homolog
 A:Reference number: A25897; MUID:86313665; PMID:3092220
 A:Accession: A25897
 A:Molecule type: mRNA
 A:Residues: 1188-1215, 1315-2224 <KA2>
 A:Cross-references: GB:M14335
 A:Note: parts of this sequence were determined by protein sequencing
 R:Keller, F.G.; Ortel, T.L.; Quinn-Alleen, M.A.; Kane, W.H.
 Biochemistry 34, 4118-4124, 1995
 A:Title: Thrombin-catalyzed activation of recombinant human factor V.
 A:Reference number: A36139; MUID:95210278; PMID:7696276
 A:Contents: annotation; thrombin cleavage sites
 C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
 C:Genetics:
 A:Gene: GDB:F5
 A:Cross-references: GDB:119896; OMIM:227400
 A:Map position: 1q23-1q23
 A:Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65
 C:Function:
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
 C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
 F:1-38/Domain: signal sequence #status predicted <SIG>
 F:29-2224/Product: coagulation factor V #status predicted <WAT>
 F:29-737/Product: coagulation factor Va heavy chain #status experimental <VAH>
 F:29-345/Domain: A1 <DA1>
 F:33-329/Domain: ferroxidase repeat homology <FO1>
 F:346-691/Domain: A2 <DA2>
 F:351-684/Domain: ferroxidase repeat homology <FO2>
 F:692-1573/Domain: B <DOB>
 F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
 F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
 F:1581-1905/Domain: A3 <DA3>
 F:1581-1905/Domain: ferroxidase repeat homology <FO3>
 F:1667-1765/Region: phospholipid binding #status predicted
 F:1906-2064/Domain: C1 <DC1>
 F:1906-2061/Domain: discoidin I amino-terminal homology <DNI>

F:2065-2224/Domain: C2 <DC2>
 F:2065-2221/Domain: discoidin I amino-terminal homology <DN2>
 F:51-55-239-297-460-468-554-741-752-760-776-782-821-938-977-1074-1083-1103-1106-1479
 F:167-193-248-329-500-526-603-684-1725-1751-1907-2061-2066-2221/disulfide bonds: #st
 F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
 F:363-693-1546/Binding site: sulfate (Tyr) (covalent) #status predicted
 F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
 F:382-1338/Binding site: carboxylate (Asn) (covalent) #status absent
 F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted
 F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimen
 F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experim
 F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental
 Alignment Scores:
 Pred. No.: 1-25e-18 Length: 2224
 Score: 351.00 Matches: 77
 Percent Similarity: 59.04% Conservative: 21
 Best Local Similarity: 46.39% Mismatches: 58
 Query Match: 9.86% Indels: 10
 Dbs: 1 Gaps: 3
 US-10-060-830-1113 (1-1962) x KFHU5 (1-2224)
 Qy 496 AAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATCGCGGATCTCTCAA 555
 Db 2062 GluValAsnGlyCysSerThrProLeuGlyMetGluAsnGlyLysIleGluAsnLysGln 2081
 Qy 556 ATACAGCATCATCTGCTGCTGGAG-----TGGACTGACACACAGGGCAAGAACAGT 609
 Db 2082 IleThrAlaSerSerPheLysLysSerTrpTrpGlyAspTyr----- 2095
 Qy 610 TGAACACCAAAAGCCAGGCTGAAAAACCTGGA-----CCGCTTGGGCTGCTTT 663
 Db 2096 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2115
 Qy 664 GCACGTGATGAATACACAGTGGTGTACAAATAGATTTTGAATAGAAAGAAATAACAGGC 723
 Db 2116 AlaAsnAsnLysGlnTrpLeuGluIleAspLeuLysIleLysLysIleThrAla 2135
 Qy 724 ATTATACACATGGATCCACCATGGTGGAGACAAATCTATGTCTGCTCCCTACAGAATC 783
 Db 2136 IleIleThrGlnGlyCysLysSerLeuSerSerGluMetTyrValLysSerTyrThrIle 2155
 Qy 784 CTGTACATGATGATGGCGAGAAATGGAGTGTGTACAGAGACCGCTGGTGGAGCAAGAT 843
 Db 2156 HisTyrSerGluGlnGlyValGluTrpLysProTyrArgLeuLysSerSerMetValasp 2175
 Qy 844 AAGATATTTCAGAGAAACAAAGATATATACAGGATGTCGTAATAACTTTTGGCCACCA 903
 Db 2176 LysIlePheGluGlyAsnThrAsnThrLysGlyHisValLysAsnPhePheAsnProPro 2195
 Qy 904 ATTATTCACGTTTATTATAGATGAATCTACCCCAATGGCAGCAAAAATGCCATGAAA 963
 Db 2196 IleIleSerArgPheIleArgValIleProLysThrTrpAsnGlnSerIleAlaLeuArg 2215
 Qy 964 ATGGAGCTGCTCGGATGT 981
 Db 2216 LeuGluLeuPheGlyCys 2221
 RESULT 4
 S55138
 glycoprotein antigen MGp57/53, mammary gland - bovine (fragment)
 N:Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
 C:Accession: S65138; G48394
 R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
 Biochim. Biophys. Acta 1245, 385-391, 1995
 A:Title: Molecular cloning of glycoprotein antigens MGp57/53 recognized by monoclonal
 A:Reference number: S65138; MUID:96125736; PMID:8541316
 A:Accession: S65138
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA

456	ATCTGT-----GGTGGGCACACTTATCTACAACTCTTTTACATTTTAAAGACAAGTGGATG	509
	::::: :::::	
Db	251 ileCysHisArgGlyCysThrLeuArgPheGlu-LeuLeuGlyCysGluLeuAsnGlyCy	270
	::::: :::::	
Qy	510 TTATGGAACACTGGGATGGAGTCTGGTGTGATCGCGGATCTCTCAATTAACAGCATCATC	569
	::::: :::::	
Db	270 sThrGluProLeuGlyLeuLysAspAsnThrIleProAsnLysGlnIleThrAlaSerSe	290
	::::: :::::	
Qy	570 TGTGCTGGAG---TGGACTGACCACACACAGGG---CAAGAGACAGATTGGAAACCCCAAAA	623
	:::::	
Db	290 rTyrTyrLysThrTrp-----GlyLeuSerAlaPheSerTrpPheProTyrTy	306
Qy	624 AGCCAGAGCTGAAAAAACCCTGGACCG-----CCTTGGGCTGCTTTTGGCCACTGATCAATA	677
Db	306 rAlaArgLeuAspAsnGlnGlyLysPheAsnAlaIleTrpThrAlaGlnThrAsnSerAlaSe	326
Qy	678 CCAGTGGTTCACAAATGATTTGAATTAAGGAAAAAGAAATAACAGGCATATTAACCACTGG	737
	:::::	
Db	326 rGluTrpLeuGlnIleAspLeuGlySerGlnLysArgValThrGlyIleIleThrGlnG	346
	:::::	
Qy	738 ATCCACCATTGGTGAGACACAATTACATATGTCTGCTGCTACAGAACTCCTGTACAGTGTATCA	797
	:::::	
Db	346 yAlaArgAspPheGlyHisIleGlnIyrValAlaIleTyrArgValAlaTyrGlyAspAs	366
	:::::	
Qy	798 TGGGCAGAAATGGACTGTGTACAGAGACCTGGTGTGGAGAGATAAGATATTTCAAGG	857
	:::::	
Db	366 pGlyValThrTrpThrGluTyrLysAspProGlyAlaSerGluSerLysIlePheProG	386
	:::::	
Qy	858 AAACAAAGATTATCACCAGGATGTGGGTATTAACACTTTTGGCCACCATAATTATTCACGTTT	917
	:::::	

Q7 100GGGHHHKKKRRRYYVGR
720 100GGGHHHKKKRRYVGR
Db pGLyValThrTriphrLysAspProGIAlaserGlusterLysilePheProGI 386
Qv 858 AAACAAAGATTATCACCAGGATGTGCCTAATAACTTTTTGCCACCAATTATTGCACGTTT 917

```

||||| ||| ... :||| ||| ||| ||| ||| |||
Db 386 yAsnMetAspAsnSerHisLysLysAsnIlePheGluThrProPheGlnAlaArgph 406

QY 918 TATTAGAGTGAATCTACCAATGGCAGACAAAATGGCATGAAATGGAGCTGCTCGG 977
|||||: ||| ||| :||| :|||: ||| ||| ||| ||| |||
Db 406 eValArgIleGlnProValAlaTrpHisAsnArgIleThrLeuArgValGluLeuGlu 426

QY 978 ATGT 981
||||
Db 426 yCys 427

RESULT 6
A47285
milk fat globule protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
C:Accession: A47285
R:Iarococa, D.; Peterson, J.A.; Urrea, R.; Kuniyoshi, J.; Blistrain, A.M.;
Cancer Res. 51, 4994-4998, 1991
A:Title: A Mr 46,000 human milk fat globule protein that is highly expressed
A:Reference number: A47285; MUID:91371351; PMID:1909932
A:Accession: A47285
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <L>AR>
A:Cross-references: GB:S56151; NID:q235396; PIDN:AAB19771.1; PID:q235397
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homolog
F:1-56/Domain: discoidin I amino-terminal homology (fragment) <DN1>
F:60-218/Domain: discoidin I amino-terminal homology <DN2>

Alignment Scores:
Pred. No.: 3,82e-18 Length: 218
Score: 343.00 Matches: 71
Percent Similarity: 57.23% Conservative: 28
Best Local Similarity: 41.04% Mismatches: 68
Query Match: 9.64% Indels: 6
DB: Gaps: 3

US-10-060-830-1113 (1-1962) x A47285 (1-218)

QY 472 TTATCTACAGTCTTTTACATTTAGACAAGTGGATGTTATGGAACACTGGGATGGAG 531
||| ||| :||| :||| ||| ||| ||| ||| |||
Db 49 LeuArgPheGluLeuLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 68

QY 532 TCTGCTGTATCGCGATCCTCAATAACAGCATCATCTGCTGGAG---TGGACTGAC 588
: ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerTyrLysThrTrpGlyLeu 88

QY 589 CACACAGGCGACAGACAGATTGGAAACCCAAAAGCCAGGCTGAAAACCTGGA--- 645
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 HisLeu-----PheSerTrpAsnProSerTyrAlaArgLeuAspLysGlnGlyAsn 105

QY 646 ---CCGCTTGGCTGCTTTTGGCACTGATGAATACCACTGGTTACAAATGATTGAAT 702
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 PheAsnAlaTrpValAlaGlySerTyrGlyAsnAspGlnTrpLeuGlnValAspLeuGly 125

QY 703 AAGGAAAAGAAAATAACAGCAGCATTAACCACTGGATCCACCATGGTGAGCACAAATTAC 762
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 SerSerLysGluValThrGlyIleIleThrGlnGlyAlaArgAsnPheGlySerValGln 145

QY 763 TATGTGTCCTCCACAGAAATCCTGTACAGTGATGATGGCAGAAATGGACTGTACAGA 822
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 146 PheValAlaSerTyrLysValAlaTyrSerAsnAspSerAlaAsnTrpThrGluTyrGln 165

QY 823 GAGCTGTGTGGAGCAAGATAGATATTTCAGAGGAACAAAGATTATCACCAGGATGTG 882
: ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 AspProArgThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisSerHisLys 185

QY 883 CGTAATAACTTTTGGCCACCAATATTGTCAGCTTTTATTAGAGTGAATCCTACCCAATGG 942
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 186 LysAsnLeuPheGluThrProIleLeuAlaArgTyrValArgIleLeuProValAlaTrp 205

QY 943 CAGCAGAAAATTGGCATGAAAATGGAGCTGCTCGGATGT 981

```

Db 206 HisAsnArgIleAlaLeuArgLeuGluLeuLeuGlyCys 218
 RESULT 7
 T11743
 PP47 protein - pig (fragment)
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 R:Accession: T11743
 R:Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsud
 Biol. Reprod. 58, 1057-1064, 1998
 A:Title: Molecular cloning and characterization of P47, a novel boar sperm-
 A:Reference number: T17325; MUID:98206817; PMID:9546740
 A:Accession: T11743
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-409 <ENS>
 A:Cross-references: EMBL:Y11683; NID:92652927; PIDN:CAA72379.1; PID:926529292
 A:Experimental source: testis
 C:Function:
 C:Superfamily: may be involved in membrane remodeling and/or function as a
 F:6-40/Domain: EGF homology <EGF>
 Alignment Scores:
 Pred. No.: 6,99e-18 Length: 409
 Score: 340.00 Matches: 76
 Percent Similarity: 56.59% Conservative: 27
 Best Local Similarity: 41.76% Mismatches: 68
 Query Watch: 9.55% Indels: 11
 DB: 2 Gaps: 5
 US-10-060-830-1113 (1-1962) x T11743 (1-409)
 QY 456 ATCTGT-----GGTGGACACTTACTACAGCTCTTTTACATTAAACAGCTGGATG 509
 Db 233 ILeCysHisArgGlyCysThrLeuArgPheGlu-LeuLeuGlyCysGluLeuSerGlyCy 252
 QY 510 TTATGGAACACTGGGGATGGAGTCTGGTGTGATCGCGGATCTCAAATAACAGCATCATC 569
 Db 252 sAlaGluProLeuGlyLeuLysAspAsnThrIleProAsnLysGlnIleThrAlaSer 272
 QY 570 TGTGCTGGAG---TGGACTGACCACACAGG---CAAGAGAACAGTTGGAAACCCAAAA 623
 Db 272 rPheTyrArgThrTrp-----GlyLeuSerAlaPheSerTriPtyrProPheTy 288
 QY 624 AGCCAGCGCTGNAANAACCTGGACCG-----CCTTGGGTGCTTTTGGCCACTGATGAATA 677
 Db 288 rAlaArgAspAsnGlnGlyLysPheAsnAlaIleThrAlaGlnSerAsnSerAlase 308
 QY 678 CCAGTGGTTACAAATAGATTGAATGAAGGAAAGAAATAACAGGCATTATAACCACTGG 737
 Db 308 rGluTrpLeuGlnIleAspLeuGlySerGlnArgArgValThrGlyIleIleThrGlnGl 328
 QY 738 ATCCACCATGGTGGGACCAATTACTATGTCTGTGCTCAGCAATCTCTACAGCATGATCA 797
 Db 328 yAlaArgAspPheGlyHisIleGlnTyrValAlaAlaTyrLysValAlaTyrSerAspAs 348
 QY 798 TGGGCAGAAATGGACTGTGTACAGAGACGCTGGTGTGGGACCAAGATAAGATATTTCACAG 857
 Db 348 pGlyValSerTrpThrGluTyrArgAspGlnGlyAlaLeuLeuGluGlyLysIlePheProGl 368
 QY 858 AAACAAGATTTATCAACGAGTGTGCGTATAACTTTTGGCCACCAATATTATGCACGTTT 917
 Db 368 yAsnLeuAspAsnAsnSerHisLysLysAsnMetPheGluThrProPheLeuThrArgPh 388
 QY 918 TATTAGATCGAATCCTACCCAATGGCAGCAAAAATGGCATGAAATGGAGCTGCTCGG 977
 Db 388 eValArgIleLeuProValAlaIleThrHisAsnArgIleThrLeuArgValGluLeuLeuGl 408
 QY 978 ATGT 981
 Db 408 yCys 409

RESULT 8

JC4915
 ags protein precursor - rat
 N:Alternate names: O-acetyl-Gd3 ganglioside
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
 C:Accession: JC4915
 R:Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
 Biochem. Biophys. Res. Commun. 225, 932-938, 1996
 A:Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
 A:Reference number: JC4915; MUID:56374422; PMID:8780713
 A:Accession: JC4915
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-427 <OCU>
 A:Cross-references: DDBJ:D84068; NID:g1620006; PIDN:BAA12210.1; PID:g1620007
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:28-60/Domain: EGF homology <EG1>
 F:68-107/Domain: EGF homology <EG2>
 F:110-267/Domain: discoidin I amino-terminal homology <DN1>
 F:271-427/Domain: discoidin I amino-terminal homology <DN2>
 C:Comment: This protein is required for the O-acetylation of disialoganglioside sialic acid.
 C:Genetics:
 A:Gene: ags
 C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homology
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:28-60/Domain: EGF homology <EG1>
 F:68-107/Domain: EGF homology <EG2>
 F:110-267/Domain: discoidin I amino-terminal homology <DN1>
 F:271-427/Domain: discoidin I amino-terminal homology <DN2>
 Alignment Scores:
 Pred. No.: 2,02e-17 Length: 427
 Score: 34.00 Matches: 67
 Percent Similarity: 58.39% Conservative: 27
 Best Local Similarity: 41.61% Mismatches: 61
 Query Match: 9.38% Indels: 6
 DB: 2 Gaps: 3

US-10-060-830-1113 (1-1962) x JC4915 (1-427)

Qy 505 GGATGTTATGAACACTGGGATGGAGTCTGTGTGATCGCGGATCTCTCAATAACAGCA 564
 Db 271 GlycylSerGluProLeuGlyLeuLysAsnThrIleProAspSerGlnIleThrAla 290
 Qy 555 TCATCTGTGCTGGAGTACGACACACAGGCGGCAAGAGACAGTGGAAACCCCAAAA 624
 Db 291 SerSer-----SerTyrLysThrTrpAsnLeuArgAlaPheGlyTyrTrpIleProHisLeu 308
 Qy 625 GCCAGGCTGAAACACCTGGACCG-----CCTTGGGCTGCTTTTGGCCACTGATCAATAC 678
 Db 309 GlyArgLeuAspAsnGlnGlyLysIleAsnAlaIleThrAlaGlnSerAsnSerAlaLys 328
 Qy 679 CAGTGGTTACAAATAGATTTTGAATAAGAAAGAAATAACAGGCATTATACCACTGGA 738
 Db 329 GluTrpLeuGlnValAspLeuGlyThrGlnLysLysValThrGlyIleIleThrGlnGly 348
 Qy 739 TCCACCATGGTGGAGCACAATATCTATGTCTGCTGCTACAGAACTCTACAGTATGAT 798
 Db 349 AlaArgAspPheGlyHisIleGlnTyrValAlaSerTyrLysValAlaHisSerAspAsp 368
 Qy 799 GGGCAGAAATGACTGTCTACAGAGACCTGCTGGGAGCAAGATACATATTTCAAGGA 858
 Db 369 GlyValGlnTrpThrValIleValGluGlnGlyThr-----SerLysValPheGlnGly 386
 Qy 859 AACAAAGATTATCACCAGGATGTGCGTAATAACTTTTGGCCACCAATATTATGACGTTT 918
 Db 387 AsnLeuAspAsnAsnSerHisLysLysAsnIlePheGluLysProPheMetAlaArgTyr 406
 Qy 919 ATTAGATGAATCTACCAATGGCAGCAGAAATATGCGCATGAAATAGAGCTGCTCGGA 978
 Db 407 ValArgValLeuProLeuSerTrpHisAsnArgIleThrLeuArgLeuGluLeuLeuGly 426
 Qy 979 TGT 981
 Db 427 Cys 427

RESULT 9

A47004
 coagulation factor VIII precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 18-Jun-1999
 C:Accession: A47004
 R:Elder, B.; Lakich, D.; Gitschier, J.
 Genomics 16, 374-379, 1993
 A:Title: Sequence of the murine factor VIII CDNA.
 A:Reference number: A47004; MUID:93300511; PMID:8314577
 A:Accession: A47004
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2319 <BLD>
 A:Cross-references: GB:L05573; NID:g192456; PIDN:AAA37385.1; PID:g192457
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxi
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:23-349/Domain: ferroxiase repeat homology <FO1>
 F:402-730/Domain: ferroxiase repeat homology <FO2>
 F:1686-2006/Domain: ferroxiase repeat homology <FO3>
 F:2007-2156/Domain: discoidin I amino-terminal homology <DN1>
 F:2160-2313/Domain: discoidin I amino-terminal homology <DN2>
 Alignment Scores:
 Pred. No.: 2,05e-16 Length: 2319
 Score: 322.00 Matches: 100
 Percent Similarity: 45.32% Conservative: 50
 Best Local Similarity: 30.21% Mismatches: 116
 Query Match: 9.05% Indels: 65
 DB: 2 Gaps: 12

US-10-060-830-1113 (1-1962) x A47004 (1-2319)

Qy 48 GGAGCTGGAGCCCA---GCAAGGCAATACTGTGTCTGGGTTGCAAA-- 95
 Db 2031 GlyGlnTrpAlaProAsnLeuAlaArgLeuHisTyrSerGlySerIleAsnAlaTrpSer 2050

Qy 96 -----GAACCATCAATTGAATCAAAAGCAATGAAATCATCATCTGTCTCATGAGTGG 149
 Db 2051 ThrLysGluProPheSerTrpIle-LysValAspLeuLeuAlaProMetIleValHisG1 2070

Qy 150 AATCCATGTTTCTGGA-----CCTGGATTTTGGC 179
 Db 2070 YileLysThrGlnGlyAlaArgGlnLysPheSerSerLeuTyrIleSerGlnPheIleI 2090

Qy 180 CTCATCTCTCTTAGATACAAAGATCTAATTACTTGTGTGGACACTGCATCCAAATTT 239
 Db 2090 eMetTyrSerLeuAspGlyLysLys-----TrpLeuSerTyrGlnGlyAsn-- 2105

Qy 240 TTTGGAACCTGAGTTCAGTAAGTACTGCCAGCTGGTTGTCTGTCTCTCTTTTGTCTGAGAT 299
 Db 2106 -----SerThrGlyThrLeuMetValPhe----- 2113

Qy 300 ATCTGGAACAATTCCTCATGGATATAGAGATTCTCGCCATTGTCATCGCTGGTGTGCA 359
 Db 2114 -----PheGlyAsnValAspSerSer----- 2120

Qy 360 TGCAGGAGTACTGTCAACACAGCTTGGCGCGCAATAGTGTGTAATTAGTAAAGGTAT 419
 Db 2121 ----GlyIleLysHisAsnSerPheAsnProPhe-----IleAlaArgTyrIle 2136

Qy 420 TCCCTATTATCAAGTCTTTTGGCTAACACAGTCACTCTGTGTGGGACACTTATCTAC 479
 Db 2136 eArgLeuHisProThrHis-----SerSerIleArgSerThrLeuArgMe 2151

Qy 480 AAGTCTTTTACATTTAAGACAAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGGTGT 539
 Db 2151 tGluLeuMetGlyCysAspLeuAsnSerCysSerIleProLeuGlyMetGluSerLysVa 2171

Qy 540 GATCGCGGATCCTCAATAACAGCATCATCTGTCTGCTGAGTGGACTGACACACAGGCA 599
 Db 2171 lileSerAspThrGlnIleThrAlaSerSerTyrPhe-----ThrAsnMe 2186

QY	600	AGAGAACACTTGGAAACCCAAAAAGCCAGCTGAAAAAACCTTGA-----CCGCCTTG	653
Db	2186	tPheAlaThrTpSerProSerGlnAlaArgLeuHisLeuGlnGlyArgThrAsnAlaTr	2206
QY	654	GGCTGCTTTTGGCAGTGAATACACCTGCTTCAATAGATTGTAATAGGAAGAA	713
Db	2206	pargProGlnValasnappProLysGlnTrpLeuGlnValAspLeuGlnLysThrMetLy	2226
QY	714	AATAACAGCATTATAACACTGGTATCCACCATGCTGGAGCACAAATACTATGTGTGC	773
Db	2226	sValThrGlyIleIleThrGlnGlyValLysSerLeuPheThrSerMetPheValLysG	2246
QY	774	CTACAGATCCCTGTCACAGTATGATGGCAGAAATGGACTGTGTACAGAGAGCCTGGTG	833
Db	2246	uPheLeuIleSerSerSerGlnAspGlyHisHisThrGlnIleLeuTyraAsnGly--	2265
QY	834	GGAGCAAGATAAGATTATTTCAAGGAACAAAGAAATATCACCAGCATGTGCGTAACT	893
Db	2266	----LysValLysValPheGlnGlyAsnGlnAspSerSerThrProMetMetAsnSerLe	2284
QY	894	TTTGGCCACCAATTATGCAGCTTTTATPAGAGTGAATCCTACCAATGGCAGAGAAAT	953
Db	2284	uAspProLeuLeuThrArgTyLeuArgIleHisProGlnIleTrpGluHisGlnI	2304
QY	954	TGCATGAATAATGGAGCTGCTCGGATGTCAG	984
Db	2304	eAlaLeuArgLeuGluIleLeuGlyLysGlu	2314
RESULT 10			
A36479			
milk fat globule membrane protein - mouse			
C:Species: Mus musculus (house mouse)			
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999			
C:Accession: A36479			
R:Stubbs, J.D.; Lekutis, C.; Singer, K.L.; Bui, A.; Yuzuki, D.; Srinivasan, Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990			
A:Title: cDNA cloning of a mouse mammary epithelial cell surface protein re			
A:Reference number: A36479; MUID:91046008; PMID:2122462			
A:Accession: A36479			
A>Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 1-463 <ST>			
C:Superfamily: GB:M38337; NID:g199142; PIDN:AAA39534.1; PID:g199143			
A:Cross-references: GB:M38337; NID:g199142; PIDN:AAA39534.1; PID:g199143			
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology			
C:Keywords: membrane protein			
F:28-60/Domain: EGF homology <EG1>			
F:68-107/Domain: EGF homology <EG2>			
F:147-303/Domain: discoidin I amino-terminal homology <DN1>			
F:307-463/Domain: discoidin I amino-terminal homology <DN2>			
Alignment Scores:			
Pred. No.:	28-16	Length:	463
Score:	321.00	Matches:	63
Percent Similarity:	57.76%	Conservative:	30
Best Local Similarity:	39.13%	Mismatches:	62
Query Match:	9.02%	Indels:	6
DB:	1	Gaps:	3
US-10-060-830-1113 (1-1962) x A36479 (1-463)			
QY	505	GGATGTATGGAACACTGGGATGGAGTCTGTTGTGATCGCGGATCTCTCAATAACAGCA	564
Db	307	GlyCysLeuGluProLeuGlyLeuLysAsnAsnThrIleProAspSerGlnMetSerAla	326
QY	565	TCATCTGCTGCTGGAGTGGACTGACCACAGCGGCAAGAGACAGTGTGAAACCCAAAAA	624
Db	327	SerSer-----SertyrLysThrTrpAsnLeuAArgAlaPheGlyTrpTyProHisLeu	344
QY	625	GCAGAGCTGAAAAAACCTGGACGG-----CCTTGGCTGCTTTTGGCACTGATGAATAC	678
Db	345	GlyArgLeuAspAsnGlnGlyLysIleAsnAlaTrpThrAlaGlnSerAsnSerAlaLys	364

QY 679 CAGTGGTTACAAATAGATTGTAATAAGCAAAAATAAACAGCGATTATTAACACTGGA 738
:::|||||
Db 365 GLutPrLeuGlnValAspLeuGlyThrGlnArgGlnValThrGlyIleileThrlngly 384
:::|||||
QY 739 TCCACCATTGTGGAGCACAATCTATCTGTCTGCCTACAGAATCCTGTACAGTGATGAT 798
:::|||||
Db 385 AlaArgaspPheGlyHisIleGlnTyrrValGlusertyrLysValalaHisSerAspasp 404
:::|||||
QY 799 GGGCAGAAATGGACTGTGTACAGAGAGCCTGTGTGGAGCAAGATAAGATATTTCACAGA 858
||| ::::::::::|
Db 405 GlyValGlnTrpThrValtyrGluGluGlnGly-----SerSerLysValPheGlnGly 422
||| ::::::::::|
QY 859 AACAAAGATTATCACCGAGATGCGGTAAATACCTTTTTGCCACCAATATTATGCAGGTTT 918
||| ||| ::::::::::|
Db 423 AsnLeuAspAsnAsnSerHisLysLysAsnIlePheGluLysProPheMetAlaArgtyr 442
:::|||||
QY 919 ATTAGATGAATCTACCCCATGCGCAGCAAAAAATGTCATGMAAATGGAGCTGCTCGGA 978
:::|||||
Db 443 ValArgValLeuProValserTrpHisasnArgIlethrLeuArgGluLeuLeugly 462

QY 979 TGT 981
|||
Db 463 Cys 463

RESULT 11

EZHU

N:Coagulation factor VIII precursor [validated] - human
N:Alternate names: antihemophilic factor A; coagulation factor Viic; procoagulant c
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence.Revision 28-Aug-1985 #text.Change 08-Dec-2000
C:Accession: I54318; A00525; I58059; A23584; A26174; A43986; S63527; S66445;
R:Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A:Title: Sequence of the exon-containing regions of the human factor VIII gene.
A:Reference number: I54318; MUID:93265012; PMID:1303178
A:Accession: I54318
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1921, 'S', 1923-2351 <RES>
A:Cross-references: GB:M88648; NID:g182381; PIDN:AAA52420.1; PID:g182383
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; See
Nature 312, 330-337, 1984
A:Title: Expression of active human factor VIII from recombinant DNA clones.
A:Reference number: A00525; MUID:85061548; PMID:6438526
A:Accession: A00525
A:Molecule type: mRNA
A:Residues: 1-2351 <WOO>
A:Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179
R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.
s, D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A:Title: Molecular cloning of a cDNA encoding human antihaemophilic factor.
A:Reference number: I58059; MUID:85061550; PMID:6438528
A:Accession: I58059
A>Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-74, 'V', 76-1259, 'E', 1261-2351 <RE2>
A:Cross-references: GB:K01740; NID:g182802; PIDN:AAA52484.1; PID:g182803
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.;
B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.
DNA 4, 333-349, 1985
A:Title: Characterization of the polypeptide composition of human factor VIII:C and
A:Reference number: A23584; MUID:86081164; PMID:3935400
A:Accession: A23584
A:Molecule type: mRNA
A:Residues: 1-2351 <TRU>
A:Cross-references: GB:M14113; NID:g182817; PIDN:AAA52485.1; PID:g182818
R:Eaton, D.; Rodriguez, H.; Vohar, G.A.
Biochemistry 25, 505-512, 1986
A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleava
ity.

A:Reference number: A26174; MUID:86159740; PMID:3082357
A:Accession: A26174

A:Molecule type: protein
A:Residues: 20-36;392-399, 'X', 401-402;1668-1678;1709-1722, 'D', 1723-1725;1741-1755 <EAT>
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.
Biochemistry 31, 3315-3325, 1992
A:Title: Identification and functional importance of tyrosine sulfate residues within re
A:Reference number: A42348; MUID:92207952; PMID:1554716
A:Accession: A42348
A:Molecule type: protein
A:Residues: 20-36;356-371;392-408;582-594;1668-1669, 'X', 1671;1672-1692;1693-1708;1709-17
A:Experimental source: recombinant material from Chinese hamster ovary cells
A:Note: sequence extracted from NCBI backbone and corrected to correspond with the publi
R:Pay, P.J.; Smudzin, T.M.
J. Biol. Chem. 264, 14005-14010, 1989
A:Title: Intersubunit fluorescence energy transfer in human factor VIII.
A:Reference number: A43986; MUID:89340500; PMID:2503509
A:Accession: A43986
A:Molecule type: protein
A:Residues: 'X', 517-523;1853-1860, 'X', 1862-1864, 'X', 1866 <PAX>
R:Leyte, A.; van Schijndel, H.B.; Niehrs, C.; Hutner, W.B.; Verbeet, M.P.; Mertens, K.;
J. Biol. Chem. 266, 740-746, 1991
A:Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for
A:Reference number: A56109; MUID:91093266; PMID:1698735
A:Contents: annotation: sulfation
R:Gitschler, J.; Wood, W.I.; Goralka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar,
Nature 312, 326-330, 1984
A:Title: Characterization of the human factor VIII gene.
A:Reference number: A56196; MUID:85061547; PMID:6438525
A:Contents: annotation: introns
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
Protein Sci. 4, 740-746, 1995
A:Title: Locations of disulfide bonds and free cysteines in the heavy and light chains of
A:Reference number: A56216; MUID:95338127; PMID:7613471
A:Contents: annotation: disulfide bonds
A:Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulphydryls
R:Kjalke, M.; Hedning, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezbhan, M.
Eur. J. Biochem. 234, 773-779, 1995
A:Title: Amino acid residues 721-729 are required for full factor VIII activity.
A:Reference number: S63527; MUID:96163459; PMID:8575434
A:Accession: S63527
A:Molecule type: protein
A:Residues: 713-752;753-759 <KTA>
R:Lind, P.; Larsson, K.; Spira, J.; Sydow-Baechman, M.; Almstedt, A.; Gray, E.; Sandberg
Eur. J. Biochem. 232, 19-27, 1995
A:Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
A:Reference number: S66445; MUID:96048024; PMID:7556150
A:Accession: S66445
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1668-1685 <LIN>
C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure pr
C:Genetics:
A:Gene: GDB:F8C
A:Cross-references: GDB:119124; OMIM:306700
A:Map position: Xq28-Xq28
A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro
A:Pathway: blood coagulation
C:Superfamily: coagulation factor VIII; discooidin I amino-terminal homology; ferroxidase
C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-2351/Product: coagulation factor VIII #status experimental <MAT>
F:20-740/Product: coagulation factor VIII heavy chain #status experimental <ACH>
F:20-356/Domain: A1 <DA1>
F:23-348/Domain: ferroxidase repeat homology <FO1>
F:392-759/Domain: A2 <DA2>
F:402-730/Domain: ferroxidase repeat homology <FO2>
F:760-1667/Domain: B <DB0>
F:1668-2351/Product: coagulation factor VIIa light chain #status experimental <ACL>
F:1709-2038/Domain: A3 <DA3>
F:1716-2038/Domain: ferroxidase repeat homology <FO3>
F:2039-2191/Domain: C1 <DC1>
F:2039-2188/Domain: discooidin I amino-terminal homology <DNI>

[illegible]

F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-927/Product: A5 antigen #status predicted <A5A>
F;27-138/Domain: C1r/C1s repeat homology <C1R1>
F;147-262/Domain: C1r/C1s repeat homology <C1R2>
E;274-424/Domain: discoidin I amino-terminal homology <DN1>
E;430-584/Domain: discoidin I amino-terminal homology <DN2>
F;646-812/Domain: MAM homology <MAM>
F;861-883/Domain: transmembrane #status predicted <TMW>
F;150,261,300,523,844/Binding site: carbohydurate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 1,4e-11 Length: 927
Score: 258.00 Matches: 121
Percent Similarity: 35.35% Conservative: 89
Best Local Similarity: 20.37% Mismatches: 226
Query Match: 7.25% Indels: 158
DB: 1 Gaps: 19

US-10-060-830-1113 (1-1962) x JQ0948 (1-927)

QY	55	GGAGCCGACGAAGCAATACTGTGCTCGGGTTGCCAAATGAACCATTCAAATTGAATCA	114
Db	221	GlyProHisIleGlyArgTyrCysGlyGlnAsnThrPro-----GlyArgValArgSer	238
QY	115	AAGGCANATGAATCACATGCTGTTTCATGAGTGGAATCCATGTTCTGGACGGGATTT	174
Db	239	PheThrGlyIleLeuSerMetilePheHisThrAspSerAlaIleAlaLysGluGlyPhe	258
QY	175	TGGCGCTCATACTCTGTTATAGATAACA--	204
Db	259	PheAlaAsnPheserValValGlnSerAsnThrAspGluAspPheGlnCysLysGluAla	278
QY	205	-----GATCTAATTACTGTTGTTG----	222
Db	279	LeuGlyMetGluSerGlyGluIleHisPheAspGlnIleserValSerSerGlnTyrSer	298
QY	222	-----	222
Db	299	MetAsnTrpSerAlaGluArgSerArgLeuAsnTyrValGluAsnGlyTrpThrProGly	318
QY	223	--GACACTGCATCCAATTTTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGGTGT	279
Db	319	GluAspThrValLysGluTrpIleGlnValAspLeuGluAsn-----	332
QY	280	CTGCTTCCTTTGCTGAGATATCTCGAACAAATTCCTCATGGATATAGAGATTCTCTGCCA	339
Db	333	--LeuArgPhe-----ValSerGlyIleGlyThrGlnGlyAlaIleSerLysGluThr	349
QY	340	TGTGCGATGGCTGGTGTGCATCCAGGAGTAGTCTCAACACAGTGGCGGCCCAATCAGT	399
Db	350	LysLysLysTyrPheValLysSeryTyrLysValAspileSerSerAsnGlyGluAspTrp	369
QY	400	GTTGTAATTAGTAAAGGTATCCCATTATTGAAAAGTCTTTGGCTAACACAGTCACATCT	459
Db	370	IleThrLeuLysAspGlyAsnLysHisLeuValPheThrGlyAsnThrAspAlaThrAsp	389
QY	460	GTTGTGGGACAC---TTATCTACAAGCTCTTTTATACA-----	492
Db	390	ValValTyrArgProPheserLysProValIleThrArgPheValArgLeuArgProVal	409
QY	493	-----TTAAGACAAGTGGATGTTATGAACA-----	519
Db	410	ThrTrpGluAsnGlyIleSerLeuArgPheGluLeuTyrGlyCysLysIleThrAspTyr	429
QY	520	-----CTGGGGATGAGTCTGGTGTGATCGCGGATCCTCAATAACACAGA	564
Db	430	ProCysSerArgMetLeuGlyMetValSerGlyLeuIleSerAspSerGlnIleThrAla	449
QY	565	TCATCTGTGCTGAGTGGACTGCACCACACAGGCGCAGAGAACACAGTTGGAAACCCAAAA	624
Db	450	SerSerGlnVal-----AspArgAsnTrpValProGluLeu	461
QY	625	GCCAGGCTGAAAAAACCTGGACGGCTTGGGTGCTTTTGGCCACTGATGAATAC-----	678

|||||
Db 462 AlaArgLeuValThrSerArgSerGlyTrpAlaLeuProProSerAsnThrHisProTyr 481
Qy 679 -----CAGTGGTTACAAATAGATTGAATGAAGGAAAGAAATAACAGCATATTAACC 732
Db 482 ThrLysGluTrpLeuGlnIleAspLeuAlaGluLysIleValArgGlyValIle 501
Qy 733 ACTGGATCCACCATGGTGGACCAATACTATGTCTGCCTACAGAACTCTAGT 792
Db 502 GlnGlyLysHisLysGluAsnLysValPheMetArgLysPheLysIleGlyTyrSer 521
Qy 793 GATGATGGCAGAAATGACTGTCTACAGAGAGCGCTGGTGGAGCAAGATAAGATATT 852
Db 522 AsnAsnGlyThrGluTrpGluMetIleMetAspSerSerLysAsnLysProLysThrPhe 541
Qy 853 CAAGGAACAAGATTATCACCAGATGTGCGTAATACATTTTGCACCACCAATTATTGCA 912
Db 542 GluGlyAsnThrAsnTyrAspThrProGluLeuArgThrPhe---AlaHisIleThrThr 560
Qy 913 CGTTTATTAGAGTAATCTTACCCTACGAGCGCTGGTGGAGCAAGATAAGATATT 852
Db 561 GlyPheIleArgIleIleProGluArgAlaSerAlaSerGlyLeuAlaLeuArgLeuGlu 580
Qy 970 CTGCTCGATGTCAGTTTATTCCTAAAGTGTCTCCCAAACTTACTCAACCTCCACCT 1029
Db 581 LeuLeuGlyCysGluVal-----GluThrProThrSerIleProThrProGlu 597
Qy 1030 CCTCGGACAGCATGACCTCAAAACACTACAGCCCTCCAAAATAGCCAAAGTCGT 1089
Db 598 ValAsnGlyLysGluCysGluGlyAspLeuAlaAsnCysHisSerGlyThrAspGlu 617
Qy 1090 GCCCAAAATTTAGCGACCACTACACCTCGAGTAGCAATGAATTCCTGCACAGACA 1149
Db 618 GlyPheLysLeuThr-----ValGlyAlaThr 626
Qy 1150 GAACAACAACCTGCAGTCCTGATATCAGAAATACTACCGTAACCTCCAAATGTAACCAA 1209
Db 627 GlyGlnSerThrGluThrProThrValGluAlaSerProGluGluProAspMetThrHis 646
Qy 1210 GATGATGCGTGGTGCAGTCTTGTCCCTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 1269
Db 647 Ser----- 647
Qy 1270 CTCATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1329
Db 648 ---AspLeuAspCysLysPheGlyTrpGlySerGlnLysThrVal----- 661
Qy 1330 GACTTACCTTACTGGGACCGGCGAGTGTGGTGGAAAGAAATGAAGCAGTCTTCTCTGCA 1389
Db 662 -----CysAsnTrpGlnHisAspIleSerSerAspLeuLysTrp 674
Qy 1390 AAAGCAGTGGACCATGAGAAACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1422
Db 675 AlaValLeuAsnSerLysThrGlyProValGlnAspHisThrGlyAspGlyAsnPheIle 694
Qy 1423 TATGACGACGACGAGTAAATAC-----CTGACGTCACAGAGAA 1461
Db 695 TyrSerGluAlaAspGluArgHisGluGlyArgAlaAlaArgLeuMetSerPro---Val 713
Qy 1462 GTCACACAGTGTGCTGCGGCTGACTCTGACAGATGTCACGACCTGCTAGGAGGAATT 1521
Db 714 ValSerSerArgSerAlaHisCysLeuThrPheTrpTyrHisMetAspGlySerHis 733
Qy 1522 GTTGTACACTTCAAGAGTCTACCTTTAAACAGAGAA 1563
Db 734 ValGlyThrLeuSerIleLysLeuLysTyrGluMetGluGlu 747

RESULT 15

JC5256

adipocyte transcription factor, AEBP1 - human

C:Species: Homo sapiens (man)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: JC5256

R:Ohno, I.; Hashimoto, J.; Shimizu, K.; Takaoka, K.; Ochi, T.; Matsubara, K.; Okubo, Biochem. Biophys. Res. Commun. 228, 411-414, 1996
A:Title: A cDNA cloning of human AEBP1 from primary cultured osteoblasts and its exp
A:Reference number: JC5256; MUID:97079196; PMID:8920928
A:Accession: JC5256
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-845 <OHN>
A:Cross-references: DDBJ:D86479; NID:g1468942; PIDN:BAAL3094.1; PID:g1468943

Alignment Scores:

Pred. No.: 4,99e-09 Length: 845
Score: 224.50 Matches: 52
Percent Similarity: 47.90% Conservative: 28
Best Local Similarity: 31.14% Mismatches: 68
Query Match: 6.31% Indels: 19
DB: 2 Gaps: 4

US-10-060-830-1113 (1-1962) x JC5256 (1-845)

Qy 520 CTGGGATGGAGTCTGGTGTGATCGGGATCCTCAATAACAGCATCATCTGTGCTGGAG 579
Db 75 IleglyMetGluSerHisArgIleGluAspAsnGlnIleArgAlaSerSerMetLeu--- 93
Qy 580 TGGACTGACACACAGGCAAGAACAGTGGAAACCCCAAAAGCCAGCGCTGAAAAA 639
Db 94 -----ArgHisGlyLeuGlyAlaGlnArgGlyArgLeuAsnMet 106
Qy 640 CCT-----GGACCGCCTTGGTGTGCTTTGCCACTGAT 672
Db 107 GlnThrGlyAlaThrGluAspAspTyrThrAspGlyAlaTrpCysAlaGluAspAla 126
Qy 673 GAATACAGTGGTTACAAATAGATTGAATAAGGAAAAAGAAATAACGGCATTAACC 732
Db 127 ArgThrGlnTrpIleGluValAspThrArgArgThrThrArgPheThrGlyValIleThr 146
Qy 733 ACTGGATCCACCATGGTGGACACAAATTAATGCTGTGCTGCTACAGAACTCTGTACAGT 792
Db 147 GlnGlyArgAspSerSerIleHisAspAspPheValThrThrPhePheValGlyPheSer 166
Qy 793 GATGATGGCAGAAATGGACTGTGTACAGAGAGCGCTGGTGGAGCAAGATAAGATATT 852
Db 167 AsnAspSerGlnThrTrpValMetTyrThrAsn---GlyTyrGlu---GluMetThrPhe 184
Qy 853 CAAGGAACAAGATTATCACCAGATGTGCGTAATAACTTTTCCCAACCAATTAATGCA 912
Db 185 HisGlyAsnValAspLysAspThrProValLeuSerGluLeuProGluProValValAla 204
Qy 913 CGTTTATTAGAGTCAATCTTACCACCAATGGCAGCAGAAAAATGCCATGAAATGAGGCTG 972
Db 205 ArgPheIleArgIleTyrProLeuThrTrpAsnGlySerLeuCysMetArgLeuGluVal 224
Qy 973 CTCGATGTCAGTATTATTCCT 993
Db 225 LeuGlyCysSerValAlaPro 231

Search completed: January 21, 2003, 09:41:54

Job time : 51.5541 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 07:26:10 ; Search time 15.5569 Seconds
(without alignments)
10461.794 Million cell updates/sec

Title: US-10-060-830-1113

Perfect score: 3559

Sequence: 1 atgctctgtctctctgct.....ttttaagaatacctttga 1962

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2.1/uspro_pool/US10060830/runat_16012003_091238_26841/app_query.fasta_1.6933
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-NODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830 -CGN_1_1_71.@runat_16012003_091238_26841 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	362	10.2	2258	FA5_PIG	O9glp1 sus scrofa
2	361	10.1	2211	FA5_BOVIN	Q28107 bos taurus
3	349	9.8	427	MFGM_BOVIN	Q95114 bos taurus
4	347	9.7	2224	FA5_HUMAN	P12259 homo sapien
5	343	9.6	387	MFGM_HUMAN	O08431 homo sapien
6	340	9.6	409	MFGM_PIG	P79385 sus scrofa
7	334	9.4	427	MFGM_RAT	P70490 rattus norv
8	322	9.0	2319	FA8_MOUSE	Q06194 mus musculus
9	321	9.0	463	MFGM_MOUSE	P21956 mus musculus
10	316	8.9	2351	FA8_HUMAN	P00451 homo sapien
11	314	8.8	925	NRP2_RAT	O35276 rattus norv
12	307	8.6	931	NRP2_MOUSE	O35375 mus musculus
13	307	8.6	2133	FA8_PIG	P12263 sus scrofa
14	298	8.4	931	NRP2_HUMAN	O60462 homo sapien
15	284.5	8.0	914	NRP1_CHICK	P79795 gallus gall
16	283.5	8.0	922	NRP1_RAT	Q9qwj9 rattus norv
17	278.5	7.8	923	NRP1_MOUSE	P97333 mus musculus
18	269.5	7.6	224	XLRL_MOUSE	Q9z114 mus musculus

19	266.5	7.5	923	1	NRP1_HUMAN	O14786 homo sapien
20	261.5	7.3	224	1	XLRL_HUMAN	O15537 homo sapien
21	258	7.2	928	1	NRP1_XENLA	P28824 xenopus lae
22	256.5	7.2	280	1	XLRL_FUGRU	Q9w6r5 fugu rubrip
23	217.5	6.1	1331	1	CTA2_HUMAN	Q9uhc6 homo sapien
24	215.5	6.1	1333	1	HMCT_BOMMO	P98092 bombyx mori
25	208	5.8	1310	1	CTA4_MOUSE	Q99p47 mus musculus
26	207.5	5.8	1284	1	NRX4_DROME	Q94887 drosophila
27	206	5.8	1308	1	CTA4_HUMAN	Q9c0a0 homo sapien
28	200.5	5.6	855	1	DDR2_HUMAN	Q16832 homo sapien
29	188.5	5.3	911	1	DDR1_MOUSE	Q03146 mus musculus
30	186	5.2	913	1	DDR1_HUMAN	Q08345 h epithelia
31	184	5.2	854	1	DDR2_MOUSE	Q62371 mus musculus
32	183.5	5.2	910	1	DDR1_RAT	Q63474 rattus norv
33	183.5	5.2	1288	1	CTA3_HUMAN	Q9b276 homo sapien
34	150	4.2	1384	1	CTA1_HUMAN	P78357 homo sapien
35	148.5	4.2	867	1	SSPO_BOVIN	P98167 bos taurus
36	145	4.1	1381	1	CTA1_RAT	P97846 rattus norv
37	145	4.1	1385	1	CTA1_MOUSE	O54991 mus musculus
38	136	3.8	257	1	D1S2_DICDI	P42530 dictyostell
39	127	3.6	837	1	LDLR_RABIT	P20063 oryctolagus
40	126.5	3.6	854	1	LDLR_CRIGR	P35950 cricetus
41	123.5	3.5	1019	1	LFC_FACTR	P28175 tachyples
42	121.5	3.4	5179	1	MUC2_HUMAN	Q02817 homo sapien
43	119.5	3.4	699	1	NP14_HUMAN	Q14978 homo sapien
44	119.5	3.4	1019	1	LFC_CARRO	Q28422 carcinoscor
45	119.5	3.4	1581	1	PPRB_HUMAN	Q15648 h peroxisom

ALIGNMENTS

RESULT 1

FA5_PIG	STANDARD;	PRT; 2258 AA.
ID	FA5_PIG	
AC	O9GLP1;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Coagulation factor V precursor (Activated protein C cofactor).	
GN	F5.	
OS	Sus scrofa (Pig).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.	
OX	NCBI_TaxID=9823;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C	
RP	DOMAINS.	
RC	TISSUE=Liver;	
EX	MEDLINE=21121490; PubMed=11229814;	
RA	Graham D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,	
RA	Kim H.K.W.;	
RT	"Porcine factor V: cDNA cloning, gene mapping, three-dimensional	
RT	protein modeling of membrane binding sites and comparative anatomy of	
RT	domains."	
RL	Cell. Mol. Life Sci. 58:148-159(2001).	
CC	-!- FUNCTION: Coagulation factor V is a cofactor that participates	
CC	with factor Xa to activate prothrombin to thrombin.	
CC	-!- SUBUNIT: Factor Va is composed of a heavy chain and a light	
CC	chain, noncovalently bound. The interaction between the two chains	
CC	is calcium-dependent.	
CC	-!- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1	
CC	and C2 may be involved in membrane binding.	
CC	-!- PTM: Thrombin activates factor V proteolytically to the active	
CC	cofactor, factor Va (formation of a heavy chain at the N-	
CC	terminus and a light chain at the C-terminus).	
CC	-!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF	
CC	2 PLASTOCYANIN-LIKE REPEATS.	
CC	-!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.	
CC	-!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-	
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL; AF191308; AAG28381.1; -
 CC HSP; P12259; ICZT.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR000421; FA58_C.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
 DR PROSITE; PS01285; FA58C.1; 2.
 DR PROSITE; PS01286; FA58C.2; 2.
 KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
 KW Repeat.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 2258 COAGULATION FACTOR V.
 FT CHAIN 23 737 HEAVY CHAIN (BY SIMILARITY).
 FT PEPTIDE 738 1611 ACTIVATION PEPTIDE (CONNECTING REGION)
 (BY SIMILARITY).
 FT CHAIN 1612 2258 LIGHT CHAIN (BY SIMILARITY).
 FT DOMAIN 30 329 F5/8 TYPE A 1.
 FT DOMAIN 30 193 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 203 329 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 348 683 F5/8 TYPE A 2.
 FT DOMAIN 348 525 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 535 683 PLASTOCYANIN-LIKE 4.
 FT DOMAIN 691 1611 B.
 FT DOMAIN 1168 1539 T-L-S-P-D-L-[GS]-[HQ]-T.
 FT REPEAT 1168 1176 1.
 FT REPEAT 1177 1185 2.
 FT REPEAT 1186 1194 3.
 FT REPEAT 1195 1203 4.
 FT REPEAT 1204 1212 5.
 FT REPEAT 1213 1221 6.
 FT REPEAT 1222 1230 7.
 FT REPEAT 1231 1239 8.
 FT REPEAT 1240 1248 9.
 FT REPEAT 1249 1257 10.
 FT REPEAT 1258 1266 11.
 FT REPEAT 1267 1275 12.
 FT REPEAT 1276 1284 13.
 FT REPEAT 1285 1293 14.
 FT REPEAT 1294 1302 15.
 FT REPEAT 1303 1311 16.
 FT REPEAT 1312 1320 17.
 FT REPEAT 1321 1329 18.
 FT REPEAT 1330 1338 19.
 FT REPEAT 1339 1347 20.
 FT REPEAT 1348 1356 21.
 FT REPEAT 1357 1365 22.
 FT REPEAT 1366 1374 23.
 FT REPEAT 1375 1383 24.
 FT REPEAT 1384 1392 25.
 FT REPEAT 1393 1401 26.
 FT REPEAT 1402 1410 27.
 FT REPEAT 1411 1419 28.
 FT REPEAT 1420 1428 29.
 FT REPEAT 1429 1437 30.
 FT REPEAT 1438 1446 31.
 FT REPEAT 1447 1455 32.
 FT REPEAT 1456 1464 33.
 FT REPEAT 1465 1473 34.
 FT REPEAT 1474 1482 35.
 FT REPEAT 1483 1491 36.
 FT REPEAT 1492 1500 37.
 FT REPEAT 1501 1509 38.
 FT REPEAT 1510 1518 39.
 FT REPEAT 1519 1527 40.
 FT REPEAT 1531 1539 41.

FT DOMAIN 1616 1941 F5/8 TYPE A 3.
 FT DOMAIN 1616 1785 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 1941 1795 PLASTOCYANIN-LIKE 6.
 FT DOMAIN 1942 2095 F5/8 TYPE C 1.
 FT DOMAIN 2100 2355 F5/8 TYPE C 2.
 FT SITE 737 738 CLEAVAGE (BY THROMBIN) (BY
 SIMILARITY).
 FT SITE 1029 1030 CLEAVAGE (BY THROMBIN) (BY
 SIMILARITY).
 FT SITE 1611 1612 CLEAVAGE (BY THROMBIN) (BY
 SIMILARITY).
 FT DISULFID 167 193 PROBABLE.
 FT DISULFID 499 525 PROBABLE.
 FT DISULFID 1759 1785 BY SIMILARITY.
 FT DISULFID 1941 2095 BY SIMILARITY.
 FT DISULFID 2100 2255 BY SIMILARITY.
 FT MOD_RES 692 692 SULFATION (POTENTIAL).
 FT MOD_RES 696 696 SULFATION (POTENTIAL).
 FT MOD_RES 724 724 SULFATION (POTENTIAL).
 FT MOD_RES 726 726 SULFATION (POTENTIAL).
 FT MOD_RES 745 745 SULFATION (POTENTIAL).
 FT MOD_RES 1560 1560 SULFATION (POTENTIAL).
 FT MOD_RES 1576 1576 SULFATION (POTENTIAL).
 FT MOD_RES 1581 1581 SULFATION (POTENTIAL).
 FT MOD_RES 1584 1584 SULFATION (POTENTIAL).
 FT MOD_RES 1588 1588 SULFATION (POTENTIAL).
 FT MOD_RES 1631 1631 SULFATION (POTENTIAL).
 FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1048 1048 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1057 1057 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1066 1066 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1480 1480 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1537 1537 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1597 1597 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1886 1886 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2243 2243 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 2258 AA; 256078 MW; 9159B9E0076A2ACC CRC64;

Alignment Scores:

Pred. No.: 4.39e-19 Length: 2258
 Score: 362.00 Matches: 79
 Percent Similarity: 58.43% Conservative: 18
 Best Local Similarity: 47.59% Mismatches: 59
 Query Match: 10.17% Indels: 10
 DB: 1 Gaps: 3

US-10-060-830-1113 (1-1962) x FA5_PIG (1-2258)

QY 496 AGACAAAGTGGATGTTATGGACACCTGGGATGCTGGTGTGATCGGGATCTCAA 555

Db 2096 GluValAsnGlyCysSerThrProLeuGlyMetGlyAsnIleLysAsnGluGln 2115

QY 556 ATAACACGATCATCTGTGCTGGAG-----TGGACTGCACACACAGGCAAGAACAGT 609

FT DISULFID 167 193 PROBABLE.
 FT DISULFID 499 525 PROBABLE.
 FT DISULFID 1712 1738 PROBABLE.
 FT DISULFID 1894 2048 BY SIMILARITY.
 FT DISULFID 2053 2208 BY SIMILARITY.
 FT MOD_RES 697 701 SULFATION (POTENTIAL).
 FT MOD_RES 701 701 SULFATION (POTENTIAL).
 FT MOD_RES 730 730 SULFATION (POTENTIAL).
 FT MOD_RES 1513 1513 SULFATION (POTENTIAL).
 FT MOD_RES 1529 1529 SULFATION (POTENTIAL).
 FT MOD_RES 1537 1537 SULFATION (POTENTIAL).
 FT MOD_RES 1541 1541 SULFATION (POTENTIAL).
 FT CARBOHYD 225 225 N-LINKED (GLCNAC).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC).
 FT CARBOHYD 382 382 N-LINKED (GLCNAC).
 FT CARBOHYD 460 460 N-LINKED (GLCNAC).
 FT CARBOHYD 553 553 N-LINKED (GLCNAC).
 FT CARBOHYD 587 587 N-LINKED (GLCNAC).
 FT CARBOHYD 745 745 N-LINKED (GLCNAC).
 FT CARBOHYD 756 756 N-LINKED (GLCNAC).
 FT CARBOHYD 774 774 N-LINKED (GLCNAC).
 FT CARBOHYD 780 780 N-LINKED (GLCNAC).
 FT CARBOHYD 902 902 N-LINKED (GLCNAC).
 FT CARBOHYD 952 952 N-LINKED (GLCNAC).
 FT CARBOHYD 964 964 N-LINKED (GLCNAC).
 FT CARBOHYD 1044 1044 N-LINKED (GLCNAC).
 FT CARBOHYD 1053 1053 N-LINKED (GLCNAC).
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC).
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC).
 FT CARBOHYD 1078 1078 N-LINKED (GLCNAC).
 FT CARBOHYD 1094 1094 N-LINKED (GLCNAC).
 FT CARBOHYD 1451 1451 N-LINKED (GLCNAC).
 FT CARBOHYD 1490 1490 N-LINKED (GLCNAC).
 FT CARBOHYD 1550 1550 N-LINKED (GLCNAC).
 FT CARBOHYD 1690 1690 N-LINKED (GLCNAC).
 FT CARBOHYD 1839 1839 N-LINKED (GLCNAC).
 FT CARBOHYD 1997 1997 N-LINKED (GLCNAC).
 FT CARBOHYD 2196 2196 N-LINKED (GLCNAC).
 FT VARIANT 587 592 NTPA -> T (IN VARIANT 2).
 SQ SEQUENCE 2211 AA; 248981 MW; CBBF90B738667C45 CRC64;

Alignment Scores:
 Pred. No.: 5,18e-19 Length: 2211
 Score: 361.00 Matches: 79
 Percent Similarity: 57.83% Conservative: 17
 Best Local Similarity: 47.59% Mismatches: 60
 Query Match: 10.14% Indels: 10
 DB: 1 Gaps: 3

US-10-060-830-1113 (1-1962) x FA5_BOVIN (1-2211)
 QY 496 AAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGGTGTGATCGCGATCCTCAA 555
 Db 2049 GluValAsnGlyCysSerThrProLeuGlyMetGluSerGlyLysIleGluAsnLysGln 2068
 QY 556 ATAACAGCATCATCTGCTGTGAG-----TGGACTGACCACACAGGGCAAGACACT 509
 Db 2069 IleThrAlaSerSerPheLysLysSerThrTrp-----GlyAsnTyr 2082
 QY 610 TGAACACCAAAAGCCAGCGCTGAAAAACCTGGA-----CCGCTTGGCTGCTTTT 663
 Db 2083 TrpGluProPheLeuAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2102
 QY 664 GCACGTGATGATACACAGCTGGTTTACAAATAGATTGAATAGGAAAGAAATAACAGGC 723
 Db 2103 AlaAsnAsnAsnGlnTrpLeuGlnIleAspLeuLysIleLysLysIleThrAla 2122
 QY 724 ATTATACCATGGATCCACCATGGTGGAGACACATTAATGCTGCTGCCTACAGAAC 783
 Db 2123 IleValThrGlnGlyCysLysSerLeuSerGluMetTyrValLysSerThrIle 2142
 QY 784 CTGTACAGTGTATGGGAGAAATGGAGCTGTGTACAGAGAGCCTGTGTGGAGCAAGAT 843

Db 2143 HistySerAspGlnGlyThrAspTrpLysProTyrArgGluLysSerSerMetValAsp 2162
 QY 844 AAGATATTTCAGGAAACAAAGATTATCACCAGGATGTCGTAATAACTTTTGGCCACA 903
 Db 2163 LysIlePheGluGlyAsnAsnValArgGlyHisValLysAsnPhePheAsnProPro 2182
 QY 904 ATTATTGCACGTTTATTAGAGTGAATCCTACCAATGCGACAGCAAAATCCCATGAAA 963
 Db 2183 IleIleSerArgPheIleArgIleIleProLysThrTrpAsnGlnSerIleAlaLeuArg 2202
 QY 964 ATGGAGCTGCTCGGATGT 981
 Db 2203 LeuGluLeuPheGlyCys 2208

RESULT 3
 MFGM_BOVIN
 ID MFGM_BOVIN STANDARD; PRT: 427 AA.
 AC Q95114; Q27959; P79344;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
 DE (MGP57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
 DE SP47) (BP47) (Components 15/16).
 GN MFG-E8.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID:9913;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
 RC STRAIN-Holstein; TISSUE-Mammary gland;
 RX MEDLINE-97008954; PubMed-8856064;
 RA Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
 RA Petersen T.E.;
 RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
 RT milk fat globules.";
 RL Eur. J. Biochem. 240:628-636(1996).
 RN [2]
 RP SEQUENCE OF 18-427 FROM N.A.
 RC TISSUE-Mammary gland;
 RX MEDLINE-96125736; PubMed-8541316;
 RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
 RA Matsuda T.;
 RT "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
 RT monoclonal antibodies raised against bovine milk fat globule
 RT membrane.";
 RL Biochim. Biophys. Acta 1245:385-391(1995).
 RN [3]
 RP SEQUENCE OF 19-427 FROM N.A.
 RC TISSUE-Testis;
 RA Enslin M.A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
 RC TISSUE-Milk;
 RX MEDLINE-93250576; PubMed-8485470;
 RA Mather I.H., Banghart L.R., Lane W.S.;
 RT "The major fat-globule membrane proteins, bovine components 15/16 and
 RT guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
 RT containing epidermal growth factor-like and factor V/VIII-like
 RT sequences.";
 RL Biochem. Mol. Biol. Int. 29:545-554(1993).
 CC -|- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
 CC MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
 CC BINDING PROTEIN.
 CC -|- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
 CC LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN.
 CC -|- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
 CC -|- PTN: THE 2 O-LINKED GLYCANS CONSTITUT OF GAL, GLCNAC AND FUC, WITH

[5] PARTIAL SEQUENCE FROM N.A.
RP TISSUE-Fibroblast; PubMed=8454869;
RX MEDLINE=93203619; PubMed=8454869;
RA Shen N.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,
RA Edgington T.S.;
RT "The serine protease cofactor factor V is synthesized by
RT lymphocytes";
RL J. Immunol. 150:2992-3001(1993).
[6]
RP SULFATION.
RX MEDLINE=94264012; PubMed=8204629;
RA Pittman D.D., Tomkinson K.N., Michnick D., Selighsohn U.,
RA Kaufman R.J.;
RT "Posttranslational sulfation of factor V is required for efficient
RT thrombin cleavage and activation and for full procoagulant activity";
RL Biochemistry 33:6952-6959(1994).
[7]
RP SULFATION.
RX MEDLINE=90366699; PubMed=2168225;
RA Hortic G.L.;
RT "Sulfation of tyrosine residues in coagulation factor V";
RL Blood 76:946-952(1990).
[8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
RX MEDLINE=20052169; PubMed=10586886;
RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
RA Fuentes-Frior P.;
RT "Crystal structures of the membrane-binding C2 domain of human
RT coagulation factor V";
RL Nature 402:434-439(1999).
[9]
RP VARIANT APCR GLN-534.
RX MEDLINE=94217810; PubMed=8164741;
RA Bertina R.M., Koelman B.P.C., Koster T., Rosendaal F.R.,
RA Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
RT "Mutation in blood coagulation factor V associated with resistance to
RT activated protein C";
RL Nature 369:64-67(1994).
[10]
RP VARIANT APCR GLY-334, AND VARIANT LYS-513.
RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese";
RL Blood 91:1135-1139(1998).
[11]
RP VARIANT APCR THR-334.
RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
RT resistance to activated protein C";
RL Blood 91:1140-1144(1998).
CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC chain, noncovalently bound. The interaction between the two chains
CC is calcium-dependent.
CC -1- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
CC REPEATS.
CC -1- PTM: Thrombin activates factor V proteolytically to the active
CC cofactor, factor Va (formation of a heavy chain at the N-
CC terminus and a light chain at the C-terminus).
CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
CC -1- DISEASE: OWREN PARAHEMOPHILIA, AN HEMORRHAGIC DIATHESIS, IS DUE
CC TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
CC FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APC
CC MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
CC A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
CC IMPLANTATION.
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS: EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L32779; AAB59401.1; JOINED.
DR EMBL; L32755; AAB59401.1; JOINED.
DR EMBL; L32756; AAB59401.1; JOINED.
DR EMBL; L32757; AAB59401.1; JOINED.
DR EMBL; L32758; AAB59401.1; JOINED.
DR EMBL; L32759; AAB59401.1; JOINED.
DR EMBL; L32760; AAB59401.1; JOINED.
DR EMBL; L32761; AAB59401.1; JOINED.
DR EMBL; L32762; AAB59401.1; JOINED.
DR EMBL; L32763; AAB59401.1; JOINED.
DR EMBL; L32764; AAB59401.1; JOINED.
DR EMBL; L32765; AAB59401.1; JOINED.
DR EMBL; L32766; AAB59401.1; JOINED.
DR EMBL; L32767; AAB59401.1; JOINED.
DR EMBL; L32768; AAB59401.1; JOINED.
DR EMBL; L32769; AAB59401.1; JOINED.
DR EMBL; L32770; AAB59401.1; JOINED.
DR EMBL; L32771; AAB59401.1; JOINED.
DR EMBL; L32772; AAB59401.1; JOINED.
DR EMBL; L32773; AAB59401.1; JOINED.
DR EMBL; L32774; AAB59401.1; JOINED.
DR EMBL; L32775; AAB59401.1; JOINED.
DR EMBL; L32776; AAB59401.1; JOINED.
DR EMBL; L32777; AAB59401.1; JOINED.
DR EMBL; L32778; AAB59401.1; JOINED.
DR EMBL; M16967; AAB52424.1; JOINED.
DR EMBL; M14335; AAB59532.1; JOINED.
DR PIR; A25897; A25897.
DR PIR; A28028; A28028.
DR PDB; 1CZS; 26-NOV-99.
DR PDB; 1CZT; 26-NOV-99.
DR PDB; 1CZV; 26-NOV-99.
DR Genew; HGNC:3542; F5.
DR MIM; 227400; -.
DR MIM; 134400; -.
DR MIM; 188055; -.
DR MIM; 227310; -.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-Oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat; Polymorphism; Disease mutation; Thrombophilia; 3D-structure.
FT SIGNAL 1 28
FT CHAIN 29 2224 COAGULATION FACTOR V.
FT CHAIN 29 737 HEAVY CHAIN.
FT PEPTIDE 738 1573 ACTIVATION PEPTIDE (CONNECTING REGION).
FT CHAIN 1574 2224 LIGHT CHAIN.
FT DOMAIN 30 329 F5/8 TYPE A 1.
FT DOMAIN 30 193 PLASTOCYANIN-LIKE 1.
FT DOMAIN 203 329 PLASTOCYANIN-LIKE 2.
FT DOMAIN 348 684 F5/8 TYPE A 2.
FT DOMAIN 348 526 PLASTOCYANIN-LIKE 3.
FT DOMAIN 536 684 PLASTOCYANIN-LIKE 4.
FT DOMAIN 692 1573 B.
FT DOMAIN 895 928 2 X 17 AA TANDEM REPEATS.
FT REPEAT 895 911 1-1.
FT REPEAT 912 928 1-2.
FT SIMILAR 1135 1148 TO 14 AA REPEATS IN BOVINE FAS.
FT DOMAIN 1185 1501 35 X 9 AA APPROXIMATE TANDEM REPEATS OF

FT REPEAT 1185 1193 [TNP]-L-S-P-D-L-S-Q-T.
 FT REPEAT 1194 1202 2-1.
 FT REPEAT 1203 1211 2-2.
 FT REPEAT 1212 1220 2-3.
 FT REPEAT 1221 1229 2-4.
 FT REPEAT 1230 1238 2-5.
 FT REPEAT 1239 1247 2-6.
 FT REPEAT 1248 1256 2-7.
 FT REPEAT 1257 1265 2-8.
 FT REPEAT 1266 1274 2-9.
 FT REPEAT 1275 1283 2-10.
 FT REPEAT 1284 1292 2-11.
 FT REPEAT 1293 1301 2-12.
 FT REPEAT 1302 1310 2-13.
 FT REPEAT 1310 1310 2-14.

Alignment Scores:
 Pred. No.: 5,65e-18 Length: 2224
 Score: 347.00 Matches: 76
 Percent Similarity: 58.43% Conservative: 21
 Best Local Similarity: 45.78% Mismatches: 59
 Query Match: 9.75% Indels: 10
 DB: 1 Gaps: 3

US-10-060-830-1113 (1-1962) x FA5_HUMAN (1-2224)

QY 496 AAGACAGTGGATGTATGGAACATGGGGATGGAGTGTGGTGTGATCGCGGATCCTCAA 555
 Db 2062 LysileThrGlnGlySerThrProLeuGlyMetGluAsnGlyLysileGluAsnLysGln 2081
 QY 556 ATAACAGCATCATCTGCTGAG-----TGGACTGACCCACAGAGCGGCAAGACAGT 609
 Db 2082 IleThrAlaSerSerPheLysSerTrpTrpGlyAspTyr----- 2095
 QY 610 TGGAAACCCAAAAGCAGCGCTGAAAAAACCCTGA-----CCGCGCTGGCTGCTTTT 663
 Db 2096 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyAArgValAsnAlaTrpGlnAlaLys 2115
 QY 664 GCACATGATGATACACAGCGTGTACAAATAGATTGAATGAAGGAAAGAAATACAGGC 723
 Db 2116 AlaAsnAsnLysGlnTrpLeuGluIleAspLeuLeuLysileLysLysileThrAla 2135
 QY 724 ATATAACACACGATGCCACATGGTGGAGCACATTAATGCTGTCTGCTCCCTACAGATC 783
 Db 2136 IleIleThrGlnGlySerLeuSerSerGluMetTyrValLysSerTrpThrile 2155
 QY 784 CTGTACAGTGTATGGGCGAGAAATGGACTGTGTACAGAGCGCTGGTGGGCAAGAT 843
 Db 2156 HistySerGluGlnGlyValGluTrpLysProTyrArgLeuLysSerSerMetValAsp 2175
 QY 844 AAGATATTTCAGGAAACAAAGATTATCACCAGGATGCGTAATAACTTTTGGCCACCA 903
 Db 2176 LysilePheGluGlyAsnThrLysGlyHisValLysAsnPhePheAsnProPro 2195
 QY 904 ATATTGACGTTTATTAGATGATCTACCCATGCCACGCAAAAATGCCATGAAA 963
 Db 2196 IleIleSerArgPheIleArgValIleProLysThrTrpAsnGlnSerIleThrLeuArg 2215
 QY 964 ATGGAGCTGCTCGGATGT 981
 Db 2216 LeuGluLeuPheGlyCys 2221

RESULT 5
 ID MFGM_HUMAN STANDARD; PRT; 387 AA.
 AC Q08431.
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)
 DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
 GN MFG8.
 OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Breast, and Breast carcinoma;
 RX MEDLINE=96213908; PubMed=8639264;
 RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
 RT "Cloning and sequence analysis of human breast epithelial antigen
 BA46 reveals an RGD cell adhesion sequence presented on an epidermal
 growth factor-like domain.";
 RL DNA Cell Biol. 15:281-286(1996).
 RN [2]
 RP SEQUENCE OF 170-387 FROM N.A.
 RC TISSUE-Mammary gland;
 RX MEDLINE=91371351; PubMed=1909932;
 RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
 RA Ceriani R.L.;
 RT "A Mr 46,000 human milk fat globule protein that is highly expressed
 in human breast tumors contains factor VIII-like domains.";
 RL Cancer Res. 51:4994-4998(1991).
 RN [3]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RC TISSUE-Milk;
 RX MEDLINE=98194924; PubMed=9535276;
 RA Giuffrida M.G., Cavalletto M., Giunta C., Conti A.,
 RA Godovac-Zimmermann J.;
 RT "Isolation and characterization of full and truncated forms of human
 breast carcinoma protein BA46 from human milk fat globule membranes.";
 RL J. Protein Chem. 17:143-148(1998).
 RN [4]
 RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
 RX MEDLINE=99342076; PubMed=10411933;
 RA Haeggqvist B., Naeslund J., Sletten K., Westmark G.T., Mucchiano G.,
 RA Tjernberg L.O., Nordsted C., Engstrom U., Westmark P.;
 RT "Medin: an integral fragment of aortic smooth muscle cell-produced
 lactadherin forms the most common human amyloid.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=97403885; PubMed=9260929;
 RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
 RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
 expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
 (RGD)-dependent cell adhesion.";
 RL DNA Cell Biol. 16:861-869(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING, BINDS
 SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
 CC -1- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
 MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
 CC -1- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
 AMINO ACID 264 AND 273.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U58516; AAC50549.1; -;
 CC EMBL: S56151; AAB19771.1; -;
 CC HSSP: P08709; 1BF9.
 CC Genew: HGNC:7036; MFG8.
 CC MIM: 602281; -;
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000421; FA58_C.
 CC Pfam: PF00008; EGF; 1.
 CC DR

DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 KW Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
 FT SIGNAL 1 23
 FT CHAIN 24 387 LACTADHERIN.
 FT CHAIN 202 387 LACTADHERIN; SHORT FORM.
 FT CHAIN 268 317 MEDIN.
 FT DOMAIN 24 67 EGF-LIKE.
 FT DOMAIN 70 225 F5/8 TYPE C 1.
 FT DOMAIN 230 387 F5/8 TYPE C 2.
 FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 27 38 BY SIMILARITY.
 FT DISULFID 32 55 BY SIMILARITY.
 FT DISULFID 57 66 BY SIMILARITY.
 FT DISULFID 70 225 BY SIMILARITY.
 FT DISULFID 212 216 BY SIMILARITY.
 FT DISULFID 230 387 BY SIMILARITY.
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 387 AA; 43123 MW; 2EE6571DEC83782D CRC64;
 Alignment Scores:
 Pred. No.: 8,19e-18 Length: 387
 Score: 343.00 Matches: 71
 Percent Similarity: 57.23% Conservative: 28
 Best Local Similarity: 41.04% Mismatches: 68
 Query Match: 9.64% Indels: 6
 DB: 1 Gaps: 3
 US-10-060-830-1113 (1-1962) x MFGM_HUMAN (1-387)
 QY 472 TTATCTACAGGCTCTTTTACATTAAAGACAGTGGATGTATGAACACTGGGATGGAG 531
 DB 218 LeuArgPheGluLeuLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 237
 QY 532 TCTGGTGTGATCGCGATCCTCAATATACAGCATCATCTGTGCTGGAG---TGGACTGAC 588
 DB 238 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerTyrLysThrTrpGlyLeu 257
 QY 589 CACACAGGCCAGACAGTGGAGAACCCAAAGAGCCAGCTGAAGAAACCTGGA--- 645
 DB 258 HisLeu-----PheSerTrpAsnProSerTyrAlaArgLeuAspLysGlnGlyAsn 274
 QY 646 ---CCGCTTGGCTGCTTTTCCACTGATGAATACCACTGGTTCACATATGATTGAAT 702
 DB 275 PheAsnAlaTrpValAlaGlySerTyrGlyAsnAspGlnTrpLeuGlnValAspLeuGly 294
 QY 703 AAGGAAGAAGAAATACAGGCATTATACCACTGGATCCACCATGGTGGGACACAAATTAC 762
 DB 295 SerSerLysGluValThrGlyIleIleThrGlnGlyAlaArgAsnPheGlySerValGln 314
 QY 763 TATGTGCTGCTACAGATCCTGTACAGTGATGATGGCAGAAATGGACTGTGTACAGA 822
 DB 315 PheValAlaSerTyrLysValAlaIleThrAlaSerSerAlaAsnTrpThrGlyLeuGln 334
 QY 823 GAGCCTGGTGGACAGATAGATATTTCAAGGAACAAAGATATATCACCAGGATGG 882
 DB 335 AspProArgThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisSerHisLys 354
 QY 883 CGTAATAACTTTTGCACCAATATTATGCAGCTTTTATTAGATGATGAATCCTACCAATGG 942
 DB 355 LysAsnLeuPheGluThrProIleLeuAlaArgTyrValArgIleLeuProValAlaTrp 374
 QY 943 CAGCAGAAATGCCATGAATATGGAGCTGCTCGGATGT 981
 DB 375 HisAsnArgIleAlaLeuArgLeuGluLeuLeuGlyCys 387

RESULT 6

MFGM_PIG
 ID MFGM_PIG STANDARD; PRT; 409 AA.
 AC PF9385;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Lactadherin (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM) (Sperm surface protein SP47) (PP47).
 DE MFGE8.
 GN MFGE8.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue=Testis;
 RA Ensslin M.A.;
 RL Submitted (NOV-1997) to the EMBL/genBank/DBJ databases.
 CC -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-BINDING PROTEIN.
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND SPERMATOZOAN.
 CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 DR EMBL; Y11683; CAA72379.1; -.
 DR HSP; P00740; IEDM.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000421; FA58_C.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 KW Glycoprotein; Repeat; EGF-like domain.
 FT DOMAIN 2 41 EGF-LIKE 1.
 FT DOMAIN 44 88 EGF-LIKE 2.
 FT DOMAIN 91 247 F5/8 TYPE C 1.
 FT DOMAIN 252 409 F5/8 TYPE C 2.
 FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 6 17 BY SIMILARITY.
 FT DISULFID 11 29 BY SIMILARITY.
 FT DISULFID 31 40 BY SIMILARITY.
 FT DISULFID 91 247 BY SIMILARITY.
 FT DISULFID 234 238 BY SIMILARITY.
 FT DISULFID 252 409 BY SIMILARITY.
 FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 409 AA; 45725 MW; B0C07AF80029927A CRC64;
 Alignment Scores:
 Pred. No.: 1.38e-17 Length: 409
 Score: 340.00 Matches: 76
 Percent Similarity: 56.59% Conservative: 27
 Best Local Similarity: 41.76% Mismatches: 68
 Query Match: 9.55% Indels: 11
 DB: 1 Gaps: 5
 US-10-060-830-1113 (1-1962) x MFGM_PIG (1-409)

QY 456 ATCTGT-----GGTGGGACACTTATCTACAAGTCCTTTTACATTTTAAGACAAAGTGGATG 509
 D 233 ILCYSHISARGGLYCYSRTHLEUARGPHEGLU-LEULEUGLYCYSGLULEUSERGGLYC 252
 QY 510 TTATGGAACTGGGATGGAGTCTGTGTGATCGCGGATCCCTCAATATACAGCATCATC 569
 D 252 SALAGLUPROLEUGLYLEULYASPNTHRIIEPROASNLYSGINLEITHRALASERSE 272
 QY 570 TGTGCTGGAG---TGGACTGACACACAGGG---CAAGAGAACAGTGTGGAAACCCAAAA 623
 D 272 rPhETyArgThrTriP-----GlyLeuSerAlapheSerTriPtyrProPheTy 288
 QY 624 AGCCAGGCTGAAAAAACCCTGGACGG-----CCTTGGGCTGCTTTTCCCACTGATGAATA 677
 D 288 rAlaArgLeuAspAsnGlnGlyLysPheAsnAlaTriPThrAlaGlnSerAsnSerAlaSe 308
 QY 678 CCAGTGGCTTACAATAGATTGAATAGAGAAAGAAATACAGGCATTATACCACTGG 737
 D 308 rGlutPLeuGlnIleAspLeuGlySerGlnArgValThrGlyIleIleThrGlnG 328
 QY 738 ATCCACCATGCTGGAGCACAAATTACTATGTCTGCTCTACAGATCCTGTACAGTATGA 797
 D 328 YAlaArgAspPheGlyHisIleGlnTyValAlaAlaTyLysValAlaTySerAspAs 348
 QY 798 TGGCGAAGATGGACTGTGTACAGAGCCCTGGTGTGGCAAGATAGATATTTCAAGG 857
 D 348 pGlyValSerTriPThrGluTyArgAspGlnGlyAlaLeuGluGlyLysIlePheProG 368
 QY 858 AAACAAGATATACACAGGATGTCGTAATACTTTTTCGCCACCAATTTATGCGACGTTT 917
 D 368 YAsnLeuAspAsnSerHisLysLysAsnMetPheGluThrProPheLeuThrArgPh 388
 QY 918 TATTAGAGTGAATCCTCCCAATGGCAGCAGAAATTCGCAATGAGGCTGCTCGG 977
 D 388 eValArgIleLeuProValAlaTriPHisAsnArgIleThrLeuArgValGluLeuLeuG 408
 QY 978 ATGT 981
 D 408 YCys 409

RESULT 7

MFQM_RAT STANDARD; PRT; 427 AA.
 ID MFQM_RAT
 AC P70490;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-acetyl GD3 ganglioside synthase) (AGS) (MFGM).
 GN MFG8 OR AGS.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96374422; PubMed=8780713;
 RA Ooura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
 RT "Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.";
 RL Biochem. Biophys. Res. Commun. 225:932-938(1996).
 CC -[1-] FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO PARTICIPATE IN THE O-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.
 CC -[1-] SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -[1-] TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
 CC -[1-] SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -[1-] SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D84068; BAA12210.1; -;
 DR HSP; P00740; IEDM;
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001438; EGF-II.
 DR InterPro; IPR000421; FA58_C.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR PRINTS; PR00010; EGFBL00D.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 KW Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
 FT SIGNAL 1 22
 FT CHAIN 23 427
 FT DOMAIN 24 61
 FT DOMAIN 64 108
 FT DOMAIN 111 267
 FT DOMAIN 272 427
 FT DISULFID 28 39
 FT DISULFID 33 49
 FT DISULFID 51 60
 FT DISULFID 68 79
 FT DISULFID 73 96
 FT DISULFID 98 107
 FT DISULFID 111 267
 FT DISULFID 254 258
 FT DISULFID 272 427
 FT SITE 87 89
 FT CARBOHYD 61 61
 FT CARBOHYD 230 230
 FT CARBOHYD 280 280
 FT CARBOHYD 390 390
 SQ SEQUENCE 427 AA; 47413 MW; EA8C8631F3EE6047 CRC64;
 Alignment Scores:
 Pred. No.: 3,87e-17 Length: 427
 Score: 334.00 Matches: 67
 Percent Similarity: 58.39% Conservative: 27
 Best Local Similarity: 41.61% Mismatches: 61
 Query Match: 9.38% Indels: 6
 DB: 1 Gaps: 3

US-10-060-830-1113 (1-1962) x MFGM_RAT (1-427)

QY 505 GGATGTTATGAACACACTGGGATGGATGCTGTGTGATCGCGGATCCCTCAATATACAGCA 564
 D 271 GlycSerGluProLeuGlyLeuLysAsnAsnThrIleProAspSerGlnIleThrAla 290
 QY 565 TCATCTGTGTGGAGTGGAGTGGACACACAGGACAGAGACAGTGTGAAACCCCAAAAA 624
 D 291 SerSer-----SerTyLysThrTrpAsnLeuArgAlaPheGlyTrpTyProHisLeu 308
 QY 625 GCCAGGCTGAAAAAACCTGGACCG-----CCTTGGGCTGCTTTTCCCACTGATGAATAC 678
 D 309 GlyArgLeuAspAsnGlnGlyLysIleAsnAlaTriPThrAlaGlnSerAsnSerAlaLys 328
 QY 679 CAGTGGTTACAATAGATTTCATTAAGAAAGAAATAACAGCATATATACCACTGGA 738
 D 329 GlutPLeuGlnValAspLeuGlyThrGlnLysValThrGlyIleIleThrGlnGly 348
 QY 739 TCACCATGCTGGAGCACAAATTACTATGTCTGCTCCTCAAGAACCTCTACAGTATGAT 798
 D 349 AlaArgAspPheGlyHisIleGlnTyValAlaSerTyLysValAlaHisSerAspAsp 368
 QY 799 GGCAGAAATGGACTGTGTACAGAGACGCTGTGGAGACGCTGTGGAGATATATTTCAAGGA 858

```

Db 369 GlyValGlnTrpThrValTyrGluGluGlnGlyThr-----SerLysValPheGlnGly 386
Qy 859 AACAAAGATTATCACCAGGATGCGTAATAACATTTTGGCCACCAATATTGCGACGTTT 918
Db 387 AsnLeuAspAsnAsnSerHisLysLysAsnIlePheGluLysProPheMetAlaArgTyr 406
Qy 919 ATTAGAGTGAATCTACCCCAATGCCACAGAAATTCGCAATGAAATGGAGCTGCTCGGA 978
Db 407 ValArgValLeuProLeuSerTrpHisAsnArgIleThrLeuArgLeuGluLeuGly 426
Qy 979 TGT 981
Db 427 Cys 427

RESULT 8
FAB_MOUSE
ID FAB_MOUSE STANDARD; PRT; 2319 AA.
AC Q06194;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR F8 OR F8C
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=93300511; PubMed=8314577;
RA Elder B., Lakich D., Gitschler J.;
RT "Sequence of the murine factor VIII cDNA.";
RL Genomics 16:374-379(1993).
CC -1- FUNCTION: FACTOR VIII. ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
CC ACTIVATED FORM, FACTOR XA.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L05573; AAA37385.1; --
DR PIR; A47004; A47004.
DR HSP; P00451; 1CFG.
DR MGD; MGI:88383; F8.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS01285; FA58C.1; 2.
DR PROSITE; PS01286; FA58C.2; 2.
KW Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
KW Signal; Glycoprotein; Sulfation.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2319 COAGULATION FACTOR VIII.
FT DOMAIN 20 349 F5/8 TYPE A 1.
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT DOMAIN 207 349 PLASTOCYANIN-LIKE 2.
FT DOMAIN 399 730 F5/8 TYPE A 2.

```

```

FT DOMAIN 399 573
FT DOMAIN 583 730
FT DOMAIN 760 1640
FT DOMAIN 1683 2008
FT DOMAIN 1683 1845
FT DOMAIN 1855 2008
FT DOMAIN 2009 2156
FT DOMAIN 2161 2313
FT SITE 391 392
FT SITE 759 760
FT SITE 1678 1679
FT SITE 1324 1325
FT SITE 1640 1641
FT MOD_RES 367 367
FT MOD_RES 737 737
FT MOD_RES 738 738
FT MOD_RES 742 742
FT MOD_RES 1669 1669
FT
FT MOD_RES 1687 1687
FT DISULFID 173 199
FT DISULFID 547 573
FT DISULFID 1819 1845
FT DISULFID 2008 2156
FT DISULFID 2161 2313
FT CARBOHYD 61 61
FT CARBOHYD 233 233
FT CARBOHYD 259 259
FT CARBOHYD 423 423
FT CARBOHYD 601 601
FT CARBOHYD 880 880
FT CARBOHYD 958 958
FT CARBOHYD 1015 1015
FT CARBOHYD 1022 1022
FT CARBOHYD 1026 1026
FT CARBOHYD 1044 1044
FT CARBOHYD 1076 1076
FT CARBOHYD 1087 1087
FT CARBOHYD 1136 1136
FT CARBOHYD 1161 1161
FT CARBOHYD 1192 1192
FT CARBOHYD 1255 1255
FT CARBOHYD 1268 1268
FT CARBOHYD 1273 1273
FT CARBOHYD 1274 1274
FT CARBOHYD 1302 1302
FT CARBOHYD 1316 1316
FT CARBOHYD 1340 1340
FT CARBOHYD 1378 1378
FT CARBOHYD 1797 1797
FT CARBOHYD 2105 2105
FT CARBOHYD 2319 AA; 266148 MW; FD054DE051DB2A01 CRG64;
SQ SEQUENCE 2319 AA; 266148 MW; FD054DE051DB2A01 CRG64;

```

Alignment Scores:

```

Pred. No.: 4.04e-16 Length: 2319
Score: 322.00 Matches: 100
Percent Similarity: 45.32% Conservative: 50
Best Local Similarity: 30.21% Mismatches: 116
Query Match: 9.05% Indels: 65
DB: 1 Gaps: 12

```

US-10-060-830-1113 (1-1962) x FAB_MOUSE (1-2319)

```

Qy 48 GGAGCTGGAGCCCA---GCAAGGCAAAATACTGTGCTGGGTGCAAAAT----- 95
Db 2031 GlyGlnTrpAlaProAsnLeuAlaArgLeuHisTyrSerGlySerIleAsnAlaTrpSer 2050
Qy 96 -----GAACCATTCATTAATCAAAAGCAATGAATACATTCGTCGTCATGAGTGG 149
Db 2051 ThrLysGluProPheSerTriple-LysValAspLeuLeuAlaProMetIleValHisG1 2070
Qy 150 AATCATGTTTCTGGA-----CGGGGATTTTGGC 179

```

Db 2070 yileLysThrGlnGlyAlaArgGlnLysPheSerSerLeuTyrIleSerGlnPheIle1 2090
 Qy 180 CTCATACCTCTCTTAGATAAACAAGATCTAATTACTTGTGTGGACACATGCATCAATTT 239
 Db 2090 emetTyrSerLeuaspGlyLysLys-----TrpLeuSerTyrGlnGlyAsn-- 2105
 Qy 240 TTTGGAACCTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTGCTCTCTCTCTCTCTGAT 299
 Db 2106 -----SerThrGlyThrLeuMetValPhe----- 2113
 Qy 300 ATCTGGACAATTCCTCATGATATAGATTCCTCGCCATTCGATCGCTGGTGTGCCA 359
 Db 2114 -----PheGlyAsnValAspSerSer----- 2120
 Qy 360 TGCAGGAGTAGTGTCAACACAGCTTGGCGGCAATCATAGTCTTGAATAGTAAGGAT 419
 Db 2121 ----GlyIleLysHisAsnSerPheAsnProPhe-----IleAlaArgTyrIle 2136
 Qy 420 TCCCTATTATGAAGTCTTTGGCTAACAAGCTCACATCTGTGTGGGACACTTATCTAC 479
 Db 2136 eArgLeuHisProThrHis-----SerSerIleArgSerThrLeuArgme 2151
 Qy 480 AAGCTCTTTTACATTAAGACAAGTGGATGTATGGAACACTGGGATGGATCTGGTGT 539
 Db 2151 tGlulMetGlyCysAspLeuAsnSerCysSerIleProLeuGlyMetGluSerLysVa 2171
 Qy 540 GATCGCGGATCCTCAAAATACAGCATCATCTGTGTGGAGTGGACTGACACACAGGCA 599
 Db 2171 lileSerAspThrGlnIleThrAlaSerSerTyrPhe-----ThrAsnme 2186
 Qy 600 AGAGAACAGTTGGAAACCCCAAAAGCCAGGCTGAAAAAACCCTGGA-----CGCCCTTG 653
 Db 2186 tPheAlaThrTrpSerProSerGlnAlaArgLeuHisLeuGlnGlyArgThrAsnAlaTr 2206
 Qy 654 GCCTCTTTCCCACTGATGAATACCACTGGTGTACAAATAGATTTGAATAGGAAGAA 713
 Db 2206 pArgProGlnValAsnAspProLysGlnTrpLeuGlnValAspLeuGlnLysThrMetLy 2226
 Qy 714 AATACAGGCAATTAACCACTGATCCACCATGCTGGAGCACAAATTAATCTGTCTGCG 773
 Db 2226 svalThrGlyIleThrGlnGlyValLysSerLeuPheThrSerMetPheValLysG1 2246
 Qy 774 CTACAGATCTCTGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 833
 Db 2246 uPheLeuIleSerSerSerGlnAspGlyHisHisTrpThrGlnIleLeuTyrAsnGly-- 2265
 Qy 834 GGAGCAAGATAGATATTTCAAGGAACAACAAGATATATCACCAGGATGCGCTAATACT 893
 Db 2266 ----LysValLysValPheGlnGlyAsnGlnAspSerSerThrProMetMetAsnSerLe 2284
 Qy 894 TTTGCCCAATTAATGACCGTTTATTAGATGAATCTTACCCCAATGCGACGAGAAAT 953
 Db 2284 uAspProLeuLeuThrArgTyrLeuArgIleHisProGlnIleTrpGluHisGlnI1 2304
 Qy 954 TCCCATGAATGAGCTGCTGGATGTCAG 984
 Db 2304 eAlaLeuArgLeuGluIleLeuGlyCysGlu 2314
 RESULT 9
 ID MFGM_MOUSE STANDARD; PRT; 463 AA.
 AC P21956; P97800;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM)
 DE (Sperm surface protein SP47) (MP47).
 GN MFG8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
 RC TISSUE-Mammary gland;
 RX MEDLINE=91046008; PubMed=2122462;
 RA Stubbs J.D., Lekutis C., Singer K.L., Bui A., Yuzuki D.,
 RA Strinivasan U., Parry G.;
 RT "CDNA cloning of a mouse mammary epithelial cell surface protein
 RT reveals the existence of epidermal growth factor-like domains linked
 RT to factor VIII-like sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
 RN [2]
 RP SEQUENCE OF 23-463 FROM N.A.
 RC TISSUE-Testis;
 RA Ensslin M.A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
 CC BINDING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
 CC SPERMATOZOAN.
 CC -1- DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
 CC TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; M38337; AAA39534.1; -;
 DR EMBL; Y11684; CAA72380.1; -;
 DR PIR; A36479; A36479.
 DR HSP; P00740; 1EDM.
 DR MGI; M102768; Mfge8.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001438; EGF-II.
 DR InterPro; IPR000421; FA58_C.
 DR Pfam; PF00008; EGF_2.
 DR Pfam; PF00754; F5_F8_type_C_2.
 DR PRINTS; PR00010; EGFBL00D.
 DR SMART; SM00181; EGF_2.
 DR SMART; SM00231; FA58C_2.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01185; EGF_2; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 KW Signal; glycoprotein; Repeat; EGF-like domain; Milk.
 FT SIGNAL 1 22
 FT CHAIN 23 463
 FT DOMAIN 24 61
 FT EGF-LIKE 1.
 FT EGF-LIKE 2.
 FT F5/8 TYPE C 1.
 FT F5/8 TYPE C 2.
 FT CELL ATTACHMENT SITE (POTENTIAL).
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT S -> F (IN REF. 2).
 FT N -> D (IN REF. 1; AA SEQUENCE).
 FT ETTYNDLGEYMTAVPTAVPTDLSNLSR ->
 FT CONFLICT 35 35
 FT CONFLICT 110 147

```
FT FT CONFLICT 168 168 G (IN REF. 2).
FT FT CONFLICT 196 196 Y -> S (IN REF. 2).
FT FT CONFLICT 309 309 L -> T (IN REF. 2).
FT FT CONFLICT 395 395 E -> A (IN REF. 2).
SQ SEQUENCE 463 AA; 51465 MW; D78B6C6EFBBA724D CRC64;

Alignment Scores:
Pred. No.: 3.6e-16 Length: 463
Score: 321.00 Matches: 63
Percent Similarity: 57.76% Conservative: 30
Best Local Similarity: 39.13% Mismatches: 62
Query Match: 9.02% Indels: 6
DB: 1 Gaps: 3

US-10-060-830-1113 (1-1962) x MFGM_MOUSE (1-463)
QY 505 GGATGTTATGACACTGGGATGAGTCTGTGTGATCGGGATCCCAATACACGA 564
DB 307 GlyCysLeuGluProLeuGlyLeuLysAsnThrIleProAspSerGlnMetSerAla 326
QY 565 TCATCTGTCTGGTGGTGGTACACACAGGAGGAGAGACAGTGGGAAACCCAAAAA 624
DB 327 SerSer-----SerTyrLysThrTrpAsnLeuArgAlaPheGlyTyrProHisLeu 344
QY 625 GCCAGGCTGAAACACCTGGACCG-----CCTGGGCTGCTTTGCCACTGATGAATAC 678
DB 345 GlyArgLeuAspAsnGlnGlyLysIleAsnAlaTrpThrAlaGlnSerAsnSerAlaLys 364
QY 679 CAGTGGTTACAATAGATTGATAGGAAGAAAGAAATACAGGCAATTATACCACTGGA 738
DB 365 GluTrpLeuGlnValSerLeuGlyThrGlnArgGlnValThrGlyIleThrGlnGly 384
QY 739 TCCACCATGGTGGACACAAATACATGTCTGCTGCTACAGAAATCTGTACAGTGTATG 798
DB 385 AlaArgAspPheGlyHisIleGlnTyrValGluSerTyrLysValAlaHisSerAsp 404
QY 799 GGGCAGAAATGACTGTGTACAGAGAGCTGTGTGGAGCAAGAGAATATTTCAAGA 858
DB 405 GlyValGlnTrpThrValTyrGluGlnGly-----SerSerLysValPheGlnGly 422
QY 859 ACAAGATTATCACAGGATGTGGTAACTATTTTCCACCAATATTTCACGCTTTT 918
DB 423 AsnLeuAspAsnSerHisLysLysAsnIlePheGluLysProPheMetAlaArgTyr 442
QY 919 ATTAGAGTGAATCCCTACCAATGGCAGCAGAGAAATTCATGAAATGAGCTGCTCGGA 978
DB 443 ValArgValLeuProValSerTrpHisAsnArgIleThrLeuArgLeuGluLeuGly 462
QY 979 TGT 981
DB 463 Cys 463

RESULT 10
F88_HUMAN
ID F88_HUMAN STANDARD; PRT; 2351 AA.
AC P00451;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component)
DE (Antihemophilic factor) (AHF).
GN F8 OR F8C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86081164; PubMed=3935400;
RA Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,
RA Hartog K., Kuo C.H., Maslarsz F.R., Merryweather J.P., Najarian R.,
RA Pacht C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,
```

```
RA Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,
RA Nordfang O., Ezban M.;
RA "Characterization of the polypeptide composition of human factor
RA VIII:C and the nucleotide sequence and expression of the human kidney
RA cDNA.";
RL DNA 4:333-349(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061548; PubMed=6438526;
RA Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschler J.,
RA Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wilson K.L.,
RA Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;
RA "Expression of active human factor VIII from recombinant DNA clones.";
RL Nature 312:330-337(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061550; PubMed=6438528;
RA Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,
RA Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,
RA Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Fass D.N.,
RA Hewick R.M.;
RA "Molecular cloning of a cDNA encoding human antihaemophilic factor.";
RL Nature 312:342-347(1984).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93265012; PubMed=1303178;
RA Gitschler J., Wood W.I.;
RA "Sequence of the exon-containing regions of the human factor VIII
RA gene.";
RL Hum. Mol. Genet. 1:199-200(1992).
RN [5]
RP SEQUENCE OF 2064-2070 FROM N.A.
RA de Water N.S., Williams R., Browett P.J.;
RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SULFATION OF TYR-1699.
RX MEDLINE=91093266; PubMed=1898735;
RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P.,
RA Mertens K., van Mourik J.A.;
RA "Sulfation of Tyr1680 of human blood coagulation factor VIII is
RA essential for the interaction of factor VIII with von Willebrand
RA factor.";
RL J. Biol. Chem. 266:740-746(1991).
RN [7]
RP SULFATION.
RX MEDLINE=92207952; PubMed=1554716;
RA Pittman D.D., Wang J.H., Kaufman R.J.;
RA "Identification and functional importance of tyrosine sulfate
RA residues within recombinant factor VIII.";
RL Biochemistry 31:3313-3325(1992).
RN [8]
RP STRUCTURE BY NMR OF 2322-2343.
RX MEDLINE=95200924; PubMed=7893714;
RA Gilbert G.E., Baleja J.D.;
RA "Membrane-binding peptide from the C2 domain of factor VIII forms an
RA amphipathic structure as determined by NMR spectroscopy.";
RL Biochemistry 34:3022-3031(1995).
RN [9]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE=91221499; PubMed=1902642;
RA Gitschler J.;
RA "The molecular basis of hemophilia A.";
RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
RN [10]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE=89088506; PubMed=2491949;
RA White G.C. II, Shoemaker C.B.;
RA "Factor VIII gene and hemophilia A.";
RL Blood 73:1-12(1989).
RN [11]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE=95245332; PubMed=7728145;
RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
```

RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
 RL Hum. Mutat. 5:1-22(1995).
 RN [12]
 RA VARIANT GLN-2326;
 RX MEDLINE-86235434; PubMed-3012775;
 RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
 RT "Identification of a missense mutation in the factor VIII gene of a
 RT mild hemophiliac";
 RL Science 232:1415-1416(1986).
 RN [13]
 RA VARIANT PRO-2135.
 RX MEDLINE-88096539; PubMed-3122181;
 RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
 RT "A novel missense mutation in the factor VIII gene identified by
 RT analysis of amplified hemophilia DNA sequences.";
 RL Nucleic Acids Res. 15:9797-9805(1987).
 RN [14]
 RA VARIANT GLN-2228.
 RX MEDLINE-88191889; PubMed-2833855;
 RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
 RA Kazazian H.H.;
 RT "Nonense and missense mutations in hemophilia A: estimate of the
 RT relative mutation rate at CG dinucleotides";
 RL Am. J. Hum. Genet. 42:718-725(1988).
 RN [15]
 RA VARIANT GLY-291.
 RX MEDLINE-88220354; PubMed-2835904;
 RA Youssoufian H., Wong C., Aronis S., Platonouk H., Kazazian H.H. Jr.,
 RA Antonarakis S.E.;
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
 RT in exon 7 of the factor VIII gene";
 RL Am. J. Hum. Genet. 42:867-871(1988).
 RN [16]
 RA VARIANT CYS-1708.
 RX MEDLINE-89274393; PubMed-2499363;
 RA O'Brien D.P., Tuddenham E.G.;
 RT "Purification and characterization of factor VIII 1,689-Cys: a
 RT nonfunctional cofactor occurring in a patient with severe hemophilia
 RT A.";
 RL Blood 73:2117-2122(1989).
 RN [17]
 RA VARIANT CYS-391.
 RX MEDLINE-90001543; PubMed-2506948;
 RA Shima M., Ware J., Yoshioka A., Fukui H., Fulcher C.A.;
 RT "An arginine to cysteine amino acid substitution at a critical
 RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
 RL Blood 74:1612-1617(1989).
 RN [18]
 RA VARIANT LEU-189.
 RX MEDLINE-90057680; PubMed-2510835;
 RA Chan V., Chan T.K., Tong T.M., Todd D.;
 RT "A novel missense mutation in exon 4 of the factor VIII:C gene
 RT resulting in moderately severe hemophilia A";
 RL Blood 74:2688-2691(1989).
 RN [19]
 RA VARIANT LEU-2326.
 RX MEDLINE-89197216; PubMed-2495245;
 RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
 RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
 RT of the factor VIII gene";
 RL Hum. Genet. 81:335-338(1989).
 RN [20]
 RA VARIANT HIS-391.
 RX MEDLINE-89264602; PubMed-2498882;
 RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Fujimaki M., Hoyer L.W.;
 RT "Direct characterization of factor VIII in plasma: detection of a
 RT mutation altering a thrombin cleavage site
 RT (arginine-372->histidine).";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
 RN [21]
 RA VARIANT CYS-1708.
 RX MEDLINE-90105723; PubMed-2104766;

RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
 RT Cys) in the factor VIII gene of two unrelated patients with
 RT cross-reacting material-positive hemophilia A.";
 RL Blood 75:384-389(1990).
 RN [22]
 RA VARIANTS GLN-2228 AND LEU-2326.
 RX MEDLINE-90123183; PubMed-2105106;
 RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
 RA Mancuso G., Morfini M., de Biasi R., Bauda F., Carbonara A.;
 RT "Recurrent mutations and three novel rearrangements in the factor
 RT VIII gene of hemophilia A patients of Italian descent.";
 RL Blood 75:662-670(1990).
 RN [23]
 RA VARIANT CYS-391.
 RX MEDLINE-90329422; PubMed-1973901;
 RA Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
 RT "CRM+ hemophilia A due to a missense mutation (372->Cys) at the
 RT internal heavy chain thrombin cleavage site";
 RL Br. J. Haematol. 75:73-77(1990).
 RN [24]
 RA VARIANTS PHE-1699 AND CYS-1708.
 RX MEDLINE-90152691; PubMed-2105906;
 RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K.,
 RA Kazazian H.H., Antonarakis S.E.;
 RT "Characterization of mutations in the factor VIII gene by direct
 RT sequencing of amplified genomic DNA.";
 RL Genomics 6:65-71(1990).
 RN [25]
 RA VARIANTS CYS-1728 AND ASP-1941.
 RX MEDLINE-90169988; PubMed-2106480;
 RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
 RA Kazazian H.H.;
 RT "Use of denaturing gradient gel electrophoresis to detect point
 RT mutations in the factor VIII gene";
 Alignment Scores:
 Pred. No.: 1,13e-15 Length: 2351
 Score: 316.00 Matches: 97
 Percent Similarity: 44.44% Conservative: 51
 Best Local Similarity: 29.13% Mismatches: 116
 Query Match: 8.88% Indels: 69
 DB: 1 Gaps: 12
 US-10-060-830-1113 (1-1962) x FA8_HUMAN (1-2351)
 QY 48 GGACGCTGGAGCCCA---GCAAGGCAATACTGTGGTGGGTGCAAAAT----- 95
 DB 2063 GlycylTrpAlaProLysLeuAlaArgLeuHisTyrSerGlySerIleAsnAlaTrpSer 2082
 QY 96 -----GAACCATTCATTAATCAAGCAAGCAATCAATGCTGTGTCATGAGTGG 149
 DB 2083 ThrLysGluProPheSerTrpIle-LysValAspLeuAlaProMetIleIleHisG1 2102
 QY 150 AATCCATGTTTCTGGA-----CCCGGATTTTGGC 179
 DB 2102 YileLysThrGlnGlyAlaArgGlnLysPheSerSerLeuTyrIleSerGlnPheIle1 2122
 QY 180 CTATATCTGTTATAGATAAACAAGATCAATTACTTGTGTTGGACACTGCATCAATTT 239
 DB 2122 eMetTyrSerLeuAspGlyLys----- 2129
 QY 240 TTTTGAACCTGAGTTCAGTAGTAC-----TGCCAGCTGGTGTCTGCTTCTCTTTC 293
 DB 2130 -----LysTrpGlnThrTyrArgGlyAsnSerThrGlyThrLeuMetValPhe-- 2145
 QY 294 TGAGATATCTGGAACAATTCCTCATGATATAGAGATTCCTCGCATTCGTCATGGCTGG 353
 DB 2146 -----PheGlyAsnValAspSerSer----- 2152
 QY 354 TGTGCATGTCAGGAGTAGTGTCAACACAGCTTGGCGGCCCAATGCTGTGTGAATAGTAA 413
 DB 354 TGTGCATGTCAGGAGTAGTGTCAACACAGCTTGGCGGCCCAATGCTGTGTGAATAGTAA 413

DE receptor 2).

GN NRP2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RC STRAIN=BA1B/C;

RX MEDLINE=97470888; PubMed=9331348;

RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;

RT "Neuropilin-2, a novel member of the neuropilin family, is a high

RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema

RT III";

RL Neuron 19:547-559(1997).

CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165

CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.

CC -1- SUBUNIT: NEUROFILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH

CC NEUROFILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN E RECEPTOR.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS; A22 (SHOWN HERE), A0, A5, A17,

CC B0 AND B5; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING CNS, PNS AND IN SOME

CC NONNEURAL TISSUES INCLUDING LIMB BUDS, DEVELOPING BONES, MUSCLES,

CC INTESTINAL EPITHELIUM, KIDNEY, LUNG AND SUBMANDIBULAR GLAND.

CC -1- DEVELOPMENTAL STAGE: THE EXPRESSION PATTERN IS VERY DYNAMIC AND

CC IS DEVELOPMENTALLY REGULATED.

CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.

CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF022856; AAC53379.1; .

DR EMBL; AF022854; AAC53377.1; .

DR EMBL; AF022855; AAC53378.1; .

DR EMBL; AF022857; AAC53380.1; .

DR EMBL; AF022858; AAC53381.1; .

DR EMBL; AF022861; AAC53382.1; .

DR HSP; P12259; ICZT.

DR MGD; MGI:1100492; Nrp2.

DR InterPro; IPR000859; CUB_domain.

DR InterPro; IPR000421; FA58_C.

DR InterPro; IPR000998; MAM_domain.

DR Pfam; PF00431; CUB; 2.

DR Pfam; PF00629; MAM; 1.

DR Pfam; PF00754; F5_F8_type_C; 2.

DR SMART; SM00042; CUB; 2.

DR SMART; SM00231; FA58C; 2.

DR SMART; SM00137; MAM; 1.

DR PROSITE; PS01180; CUB; 2.

DR PROSITE; PS01285; FA58C_1; 2.

DR PROSITE; PS01286; FA58C_2; 2.

DR PROSITE; PS00060; MAM_2; 1.

KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;

KW Alternative splicing.

FT SIGNAL 1 20

FT CHAIN 21 931

FT DOMAIN 21 864

FT TRANSMEM 865 889

FT DOMAIN 890 931

FT DOMAIN 28 142

FT DOMAIN 149 267

FT DOMAIN 277 427

FT DOMAIN 434 592

FT DOMAIN 642 802

FT POTENTIAL.

FT NEUROFILIN-2.

FT EXTRACELLULAR (POTENTIAL).

FT POTENTIAL.

FT CYTOPLASMIC (POTENTIAL).

FT CUB 1.

FT CUB 2.

FT F5/8 TYPE C 1.

FT F5/8 TYPE C 2.

FT MAM.

FT	DOMAIN	838	845	
FT	DISULFID	28	55	POLY-SER.
FT	BY SIMILARITY.			
FT	DISULFID	83	105	BY SIMILARITY.
FT	DISULFID	149	175	BY SIMILARITY.
FT	DISULFID	208	230	BY SIMILARITY.
FT	DISULFID	277	427	BY SIMILARITY.
FT	DISULFID	434	592	BY SIMILARITY.
FT	CARBOHYD	152	152	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	157	157	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	839	839	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	809	813	MISSING (IN ISOFORM A17).
FT	VARSPLIC	809	830	MISSING (IN ISOFORM A0).
FT	VARSPLIC	814	830	MISSING (IN ISOFORM A5).
FT	VARSPLIC	810	931	EDPKVDIPETHGEGEYDEIDDEYEGDWSNSSSTSGAGDP
FT				SSGKESWLITLPILITIIAMSLGVLGATCAGLLIYCT
FT				CSYSGLSRSCCTLENTFELYDGLKHVKINHOKCSEA
FT				-> GTPPGTEPTVTPVQPIPAYIYVMAAGAVLVLAS
FT				VVLALVHLHFRFYAAKTDHSTYKTSHTYNGAPLAVEPT
FT				LTIKLEOERGSCH (IN ISOFORM B0).
FT	VARSPLIC	814	931	VDIPEHGGEGYEIDDEYEGDWSNSSSTSGAGDPSSGK
FT				EKSWLYTLDPIILITIIAMSLGVLGATCAGLLIYCTSY
FT				GLSSRSCCTLENTFELYDGLKHVKINHOKCSEA -> G
FT				GTPPGTEPTVTPVQPIPAYIYVMAAGAVLVLASVVL
FT				ALVHLHFRFYAAKTDHSTYKTSHTYNGAPLAVEPTLT
FT				KLEOERGSCH (IN ISOFORM B5).
FT	CONFLICT	786	786	G -> I (IN REF. 1: AAC53380/AAC53381).
FT	SEQUENCE	931 AA;	104558 MW;	76F2443F411D2F63 CRC64;
Alignment Scores:				
Pred. No.:	4.44e-15	Length:	931	
Score:	307.00	Matches:	176	
Percent Similarity:	34.92%	Conservative:	88	
Best Local Similarity:	23.28%	Mismatches:	287	
Query Match:	8.63%	Indels:	205	
DB:	1	Gaps:	33	
US-10-060-830-1113 (1-1962) x NRP2_MOUSE (1-931)				
Qy	67	GGCAATACTGCTGGTGGGTTGCAATGAAC-----CATTCATTAATGAATCAAAA	117	
Db	102	GlyLysHisCysGly-----AsnileAlaProThrIleSer	116	
Qy	118	GGCAATGAATACACATTCCTGCTCATGAGTGAATCCATGTTCTGGACGGGATTTTG	177	
Db	117	GlySerValLeuTyrIleLeuPheThrSerAspTyrAlaArgGlnGlyAlaGlyPheSer	136	
Qy	178	GCCTCATACTCTGTTATAGATAAACAAGATCTAATTAATCTTTGGACACTGCATCCAAT	237	
Db	137	LeuArgTyrGluIlePheLeuThrGlySer-----GluAspCysSerLysAsn	152	
Qy	238	TTTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTGCTCTCTTTTGTCTGAG	297	
Db	153	PheThrSerProAsnGlyThrIleGluSerProGly-----PheProGlu	167	
Qy	298	ATATCTGGAACAATTCCTCATGGATATAGATATCCCTGCCCATCTGTCATGGCTGGTGTG	357	
Db	168	-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAlaLysPro	183	
Qy	358	CATGCAGGAGTAGTGTCACACACAGTTGGCGGCCCAATACGT-----	399	
Db	184	ArgMetGluIleLeuLeuGlnPheLeuThrPheAspLeuGluHisAspProLeuGlnVal	203	
Qy	400	-----GTTGTAATTAGTAAGGTATTCCC-----	423	
Db	204	GlyGluGlyAspCysLysTyrAspTrpLeuAspIleTpaSpGlyIleProHisValGly	223	
Qy	424	-----TATTATGAAGTCTTGGCTACACAGTCACATCTGTGGTGGGA	468	
Db	224	ProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSerThrGly	243	
Qy	469	CACATTACTACAAGCTCTTTTACATTTAAGACAAGT-----	504	

Alignment Scores:	
Pred. No.:	4.44e-15
Score:	307.00
Percent Similarity:	34.92%
Best Local Similarity:	23.28%
Query Match:	8.63%
DB:	1

Db 244 IleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAspGlyPhe 260
QY 505 -----GGATGTTATGGACA 519
Db 261 SerAlaArgTyrTyrLeuIleHisGlnGluProGluAsnPheGlnCysAsnValPro 280
QY 520 CTGGGATGAGTCTGTGTGATCGCGATCTCAAAATACAGCATCATCTGTCTGGAG 579
|||||
Db 281 LeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer----- 297
QY 580 TGGACTGACCACACAGGCGCAACAGATTGGAAAGCCAAAGCCAGCGTGAAGAAA 639
|||||
Db 298 -----ThrPheSerAspGlyArgTyrThrProGlnGlnSerArgLeuHisGly 313
QY 640 CTGGACCGCGCTGGCGTCTTTTCCACCTGATGATACACAGTGGTTACAAATAGATTG 699
|||||
Db 314 AspAspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnValAspLeu 333
QY 700 AATAAGGAAAGAAATAACAGGCATTATTAACCACTGGA-----TCCACCATGGTGGAG 753
|||||
Db 334 ArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGluThrGln 353
QY 754 CACAATTACTATGTCCTCAGACATCTGTACAGATGATGGCAGCAAAATGGACT 813
|||||
Db 354 LysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAspTyrMet 373
QY 814 GTGTACAGAGCGCTGTGTGGACAGATAGATATTTCAAGGAACAAGATTATCAC 873
|||||
Db 374 ValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAspAlaThr 391
QY 874 CAGGATGTGGTATAAATCTTTTCCCAACCAATATTACGCTTTTATTAGATGAATCCT 933
|||||
Db 392 GluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIleArgPro 411
QY 934 ACCCAATGGCAGCAAAATTTGCCATGAAATGGAGTGTGTGGATGTCATGTTATTCCT 993
|||||
Db 412 GlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 994 AAGGTGCTCTCAAAACITACTACACCTCCACCTCTCGGACACGAATGACCTCAA 1053
|||||
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMetLeuSer 442
QY 1054 AACACTACAGCCCTCCAAAATAGCAAGGTGCTGCCCAAAATTT----- 1101
|||||
Db 443 GlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTyrSerPro 462
QY 1102 ---ACGCAACCCATACACCTCGGAGTAGCAATGAATTCCTGCACAGACAGACAACA 1158
|||||
Db 463 SerAlaAlaArgLeuValSerSerArgSerGlyTyrPheProArgAsnProGlnAlaGln 482
QY 1159 ACTGCCAGTCTGATATCAGAAATACACCTTAACCTCAATGTAACCAAGATGTA--- 1215
|||||
Db 483 ProGlyGluGluTyrLeuGlnValAspLeuGlyThrProLysThrValLysGlyValle 502
QY 1216 -----CGCGTGTGCTGCAGTTCTTGTCCCTGCTGCTGCTGCTGCTGCTGCTG 1251
|||||
Db 503 IleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPheValArg 522
QY 1252 GTCCTCACTACTCCTCATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1296
|||||
Db 523 LysPheLysValSerTyrSerLeuAsnGlyLysAspTyrGluTyrIleGlnAspProArg 542
QY 1297 AGAAACAGAAAGAAACCAAGGACCACTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1350
|||||
Db 543 ThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1351 GCAGGTGGTGAAGGAATGAAGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAA 1410
|||||
Db 559 -----AspIleArgArgPheAspPro----- 565
QY 1411 ACCCCAGTTCGCTATAGCAGGAGGCAAGTTAATCACCCTGATGTCACCAAGAGATCACCACA 1470
|||||
Db 566 ValProAlaGlnTyrValArgValTyrProGluArgTyrSerProAlaGlyIleGlyMet 585

QY 1471 GTGCTGCGAGCTGACTCTGCAGAGTAT-----GCTCAGCCACCTGGTAGGAGTAATGTT 1524
|||||
Db 586 ArgLeuGluValLeuGlyCysAspTyrThrAspSerLysProThrValGluThrLeuGly 605
QY 1525 GGTACACTTCAATCAAGATCTACCTTTAAACCA-----GAAGAAGGAAAGAA 1572
|||||
Db 606 ProThrValLysSerGluGluThrThrProTyrProMetAspGluAspAlaThrGlu 625
QY 1573 GCAGGCTAT-----GCAGACCTAGATCTCTCACTACACTCACCAGG 1611
|||||
Db 626 CysGlyGluAsnCysSerPheGluAspLysLeuGln-LeuProSerGlyPheAs 645
QY 1612 CAGGAAGTTTATCATGCT-----ATGCTGAACCACTCCCA---TT 1650
|||||
Db 645 nCysAsnPheAspPheProGluGluThrCysGlyTyrValTyrAspHisAlaLysTrpLe 665
QY 1651 ACGGGCGCTGAGTATGCAACCCCAATCATCATGACATGTCCAGGCGACCCCA----- 1702
|||||
Db 665 uArg-----SerThrTyrIleSerSerAlaAsnProAsnAspAr 678
QY 1703 -----CAACTTCAG----- 1711
Db 678 gThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArgGluGlyG 698
QY 1712 -----TTGCTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCACTCCC 1761
|||||
Db 698 nTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProArgSerP 713
QY 1762 CCACTAGTGGGAACCTTACAATACACTTCTCTC----- 1793
|||||
Db 713 roValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuGlnValV 733
QY 1794 -----CAGACTGACAGCTGCTC-----CTCAGCCCGAGCCCGCAGTATGAT 1833
Db 733 alArgGluAlaSerGlnGlnSerLysLeuLeuTyrValIle-ArgGluAspGlnGlySer 752
QY 1834 ACCCGAAAGCTGGGAAGCCAGGCTACTGCGCCAGACGAATGTTGGTACAGGTGCCA 1893
|||||
Db 753 GluTyrLysHisGlyArgIleLeuLeuProSerTyrAsp---MetGluTyrGlnIle--- 770
QY 1894 CAGAGCACACAAAGATATCAGAGCAGGAAGGAGGATGGGAA 1935
|||||
Db 771 ---ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu 783
RESULT 13
FAB_PIG STANDARD; PRT; 2133 AA.
AC P12263; Q95243;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor VIII precursor (procoagulant component).
GN F8 OR CF8.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [J]
RP SEQUENCE FROM N.A.
RA Healey J.F., Lubin I.M., Lollar P.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 705-1573 FROM N.A.
RX MEDLINE=86287369; PubMed=3016730;
RA Toole J.J., Pittman D.D., Orr E.C., Murtha P., Wasley L.C.,
RA Kaufman R.J.;
RT "A large region (approximately equal to 95 kDa) of human factor VIII
is dispensable for in vitro procoagulant activity";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5939-5942(1986).
RN [3]
RP SEQUENCE OF 392-759 FROM N.A.
RX MEDLINE=94179260; PubMed=7510693;

RA Lubin I.M., Healey J.F., Scandella D., Runge M.S., Lollar P.;
 RT "Elimination of a major inhibitor epitope in factor VIII.";
 RL J. Biol. Chem. 269:8639-8641(1994).
 CC -1- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
 CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
 CC ACTIVATED FORM, FACTOR XA.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
 CC 2 PLASTOCYANIN-LIKE REPEATS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U49517; AAB06705.1; -
 DR PIR; A25945; A25945.
 DR HSSP; P00451; ICFG.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR000421; FA58_C.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 DR PROSITE; PS01285; FA58C.1; 2.
 DR PROSITE; PS01286; FA58C.2; 2.
 DR Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
 KW Signal; Glycoprotein; Sulfation.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2133 COAGULATION FACTOR VIII.
 FT DOMAIN 20 357 F5/8 TYPE A 1.
 FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 207 357 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 399 573 F5/8 TYPE A 2.
 FT DOMAIN 583 730 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 760 1599 PLASTOCYANIN-LIKE 4.
 FT DOMAIN 1495 1822 F5/8 TYPE A 3.
 FT DOMAIN 1495 1659 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 1669 1822 PLASTOCYANIN-LIKE 6.
 FT DOMAIN 1823 1970 F5/8 TYPE C 1.
 FT DOMAIN 1975 2127 F5/8 TYPE C 2.
 FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 1449 1450 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
 FT SITE 1490 1491 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT MOD_RES 737 737 SULFATION (BY SIMILARITY).
 FT MOD_RES 738 738 SULFATION (BY SIMILARITY).
 FT DISULFID 173 199 PROBABLE.
 FT DISULFID 547 573 PROBABLE.
 FT DISULFID 1633 1659 PROBABLE.
 FT DISULFID 1822 1970 BY SIMILARITY.
 FT DISULFID 1975 2127 BY SIMILARITY.
 FT CARBOHYD 233 233 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 601 601 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 929 929 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 985 985 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1025 1025 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1111 1111 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1181 1181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1208 1208 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1245 1245 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1265 1265 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1335 1335 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1408 1408 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1611 1611 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 1919 1919 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 713 713 N -> M (IN REF. 2).
 FT CONFLICT 734 734 I -> T (IN REF. 2).
 FT CONFLICT 792 792 G -> Q (IN REF. 2).
 FT CONFLICT 1133 1133 E -> F (IN REF. 2).
 FT CONFLICT 1191 1191 I -> L (IN REF. 2).
 FT CONFLICT 1209 1209 R -> F (IN REF. 2).
 FT CONFLICT 1437 1437 C -> G (IN REF. 2).
 FT CONFLICT 1456 1456 F -> R (IN REF. 2).
 FT CONFLICT 1539 1539 F -> R (IN REF. 2).
 FT CONFLICT 1546 1546 Q -> N (IN REF. 2).
 SQ SEQUENCE 2133 AA; 239304 MW; 152BBA8997F570DA CRC64;
 Alignment Scores:
 Pred. No.: 5.15e-15 Length: 2133
 Score: 307.00 Matches: 110
 Percent Similarity: 38.52% Conservative: 51
 Best Local Similarity: 26.32% Mismatches: 163
 Query Match: 8.63% Indels: 94
 DB: 1 Gaps: 12
 US-10-060-830-1113 (1-1962) x FA8_PIG (1-2133)
 QY 130 ACATTGCTGTCATGAGTGAATCCATGTTCTTGGACGGGATTTTGGCTCATCTCT 189
 DB 1663 ThrLeuAsnAlaAlaHisGlyArgGlnValThrValGlnGluPheAlaLeuPheThr 1682
 QY 190 GTTATAGATAAACAGATCTAATTTACTGTTGGACACTGCATCCCAATTTTGGAACT 249
 DB 1683 IlePheAsp-----GlutThrLysSerTrpTyrPheThrGlu 1694
 QY 250 GAGTTCAGTAAGTACTGCCAGCTGGTGTGCTGCTTCTTCTGCTGATATCTGGACAA 309
 DB 1695 AsnValGluArgAsnCysArgAlaProCysHisLeuGlnMetGluAsp-----ProThr 1712
 QY 310 ATTCTCATGATATAGAGATTCTCCCATTTGTCATGCTGCTGTCATGCA----- 363
 DB 1713 LeuLysGluAsnTyrArg-----PheHisAlaIleAsn 1723
 QY 364 GGAGTAGTGTCAAACACAGTTGGCGGCAATCAGTGTGCTTAATAGTAAGGTATTCCT 423
 DB 1724 GlyTyrValMetAspThrLeuProGlyLeuVal---MetAlaGlnAsnGlnArgIleArg 1742
 QY 424 TATATGAAAGTCTTTGGCTACAAAC----- 450
 DB 1743 TrpTyrLeuLeuSerMetGlySerAsnGlnIleHisSerIleHisPheSerGlyHis 1762
 QY 450 ----- 450
 DB 1763 ValPheSerValArgLysLysGluGluTyrLysMetAlaValTyrAsnLeuTyrProGly 1782
 QY 451 -----GTCACATCTGTGTGGGA----- 468
 DB 1783 ValPheGluThrValGluMetLeuProSerLysValGlyIleTrpArgIleGluCysLeu 1802
 QY 469 -----CACTATCTACAAGTCTT-----TTTACATTTAGACAAAGTGGATGT 510
 DB 1803 IleGlyGluHisLeuGlnAlaGlyMetSerThrThrPheLeuValTyrSerLysGluCys 1822
 QY 511 TATGGACACTGGGATGGAGTGTGTGATCGCGATCTCAATAACAGCATCATCT 570
 DB 1823 GlnAlaProLeuGlyMetAlaSerGlyArgIleArgPheGlnIleThrAlaSer--- 1841
 QY 571 GTGCTGGAGTGGACTGACACAGGCGCAAGAGAACAGTGTGGAAACCCCAAAAGCCAGG 630
 DB 1842 -----GlyGlnTyrGlyGlnTrpAlaProLysLeuAlaArg 1853
 QY 631 CTGAAAAAAGCTGGACCGCTTGGCTGCTTTTCCCACTGATGATACAGGTGTTACAA 690
 DB 1854 LeuHisTyrSerGlySerIleAsnAlaTrpSerThrLysAspProHisSerTrpIleLys 1873
 QY 691 ATAGATTGATAGGAAAGAAATAACAGGCATTATAACCACTGGATCCACCATGGT 750
 DB 691 ATAGATTGATAGGAAAGAAATAACAGGCATTATAACCACTGGATCCACCATGGT 750

Db 1874 ValAspLeuLeuAlaProMetIleIleHisGlyIleMetThrGlnGlyAlaArgGlnLys 1893
 QY 751 GAGCAATTAAGTCTGCTGCCTACAGAAATCCGTGACAGTGTGAGGCGAGAAATGG 810
 Db 1894 PheSerLeuTyIleSerGlnPheIleIleMetTyIleSerLeuAspGlyArgAsnTrp 1913
 QY 811 ACTGTGTACAGACCCGTGGTGTGGAGCAAGATAGATATTCAGAGAAACAAGATTAT 870
 Db 1914 GlnSerTyArgGlyAsnSerThrGlyThrLeuMetValPhePheGlyAsnValAspAla 1933
 QY 871 CACCAGGATGTGCGTAATAACTTTTCCACCAATATATGACCGTATTATTAGATGAAT 930
 Db 1934 SerGlyIleLysHisAsnIlePheAsnProIleValAlaArgTyIleArgLeuHis 1953
 QY 931 CTACCCAAATGACGACGAGAAATCCCATGAATGAGGAGCTGCTGGATGTCAGTTT--- 987
 Db 1954 ProThrHisTyIleArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsn 1973
 QY 988 -----ATTCCFAAAGGTGCTCTCCAAACCTTACTCAACCTCCACTCCTCGGAAC 1038
 Db 1974 SerCysSerMetProLeuGlyMetGlnAsnLysAlaIleSerAspSerGlnIleThrAla 1993
 QY 1039 AGCAATGACCTCAAAAC-----ACTACAGCCCTCCAAAATAGCCAAAGGTGCTGCC 1092
 Db 1994 SerSerHisLeuSerAsnIlePheAlaThrTrpSerProSerGlnAlaArgLeuHisLeu 2013
 QY 1093 CCAAAATTTACGACACCACTACACCTCGCAGTACGATGATTTCTTCGACAGACAGA 1152
 Db 2014 GlnGlyArgThrAsnAlaTrpArgProArgValSer-----SerAla 2027
 QY 1153 CAACAACCTGCGAGTCCGATATCAGAAATACCTACCGTAACCTCCAAATGTAACC 1206
 Db 2028 GluGluTrpLeuGlnValAspLeuGlnLysThrValLysValThrGlyIleThr 2045
 RESULT 14
 NRP2_HUMAN STANDARD; PRT; 931 AA.
 AC 060462; 014820; 014821;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
 GN NRP2 OR VEGF165R2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
 RX MEDLINE=97470888; PubMed=9331348;
 RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
 RT "Neuropilin-2, a novel member of the neuropilin family, is a high
 RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
 RT III.";
 RL Neuron 19:547-559(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM A22).
 RC TISSUE=Breast;
 RX MEDLINE=98188099; PubMed=9529250;
 RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagesbrun M.;
 RT "Neuropilin-1 is expressed by endothelial and tumor cells as an
 RT isoform-specific receptor for vascular endothelial growth factor.";
 RL Cell 92:735-745(1998).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=20309748; PubMed=10748121;
 RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
 RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
 RT form of vascular endothelial growth factor (VEGF) and of placenta
 RT growth factor-2, but only neuropilin-2 functions as a receptor for
 RT the 145-amino acid form of VEGF";
 RL J. Biol. Chem. 275:18040-18045(2000).

CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
 CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
 CC -1- SUBUNIT: NEUROFILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
 CC NEUROFILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN 3C RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A22 (SHOWN HERE), A0 AND A17;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF022859; AAC51788.1; --
 CC EMBL; AF022860; AAC51789.1; --
 CC EMBL; AF016098; AAC12922.1; --
 CC HSSP; P12259; 1CZT.
 CC Genew; HGNC:8005; NRP2.
 CC MIN; 602070; --
 CC InterPro; IPR000859; CUB_domain.
 CC InterPro; IPR000421; FA58_C.
 CC InterPro; IPR000998; MAM_domain.
 CC Pfam; PF00431; CUB; 2.
 CC Pfam; PF00629; MAM; 1.
 CC Pfam; PF00754; F5_F8_type_C; 2.
 CC SMART; SM00042; CUB_2; 2.
 CC SMART; SM00231; FA58C; 2.
 CC SMART; SM00137; MAM; 1.
 CC PROSITE; PS01180; CUB; 2.
 CC PROSITE; PS01285; FA58C_1; 2.
 CC PROSITE; PS01286; FA58C_2; 2.
 CC PROSITE; PS50060; MAM_2; 1.
 CC Transmembrane; Glycoprotein; Neuropeptide; Receptor;
 CC Alternative splicing.
 CC SIGNAL 1 20
 CC CHAIN 931
 CC DOMAIN 21 864
 CC TRANSMEM 865 889
 CC DOMAIN 890 931
 CC DOMAIN 28 142
 CC DOMAIN 149 267
 CC DOMAIN 277 427
 CC DOMAIN 434 592
 CC DOMAIN 642 802
 CC DOMAIN 671 674
 CC DISULFID 28 55
 CC DISULFID 83 105
 CC DISULFID 149 175
 CC DISULFID 208 230
 CC DISULFID 277 427
 CC DISULFID 434 592
 CC CARBOHYD 152 152
 CC CARBOHYD 157 157
 CC CARBOHYD 629 629
 CC CARBOHYD 839 839
 CC CARBOHYD 809 813
 CC VARSPPLIC 809 830
 CC CONFLICT 602 602
 CC SEQUENCE 931 AA; 104830 MW; 270CBAE69A0A797C CRC64;
 Alignment Scores:
 Pred. No.: 2.06e-14 Length: 931
 Score: 298.00 Matches: 171
 Percent Similarity: 34.93% Conservative: 92
 Best Local Similarity: 22.71% Mismatches: 291
 Query Match: 8.37% Indels: 199

	DB:	1	Caps:	32	
		US-10-060-830-1113 (1-1962) x NRP2_HUMAN (1-931)			
QY	67	GGCAATACCTGCTGTGGTTCGAATGAAC	-----CATTCAATTGAATCAAAA	117	
Db	102	GlyHisCysGly	: : : :	116	
QY	118	GGCAATGAATACATCTGTTTCATGAGTGAATCCATGTTCTGGACGGGATTTTTG	: : : :	177	
Db	117	GlySerMetLeuValIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGlyPheSer	: : : :	136	
QY	178	GCCTCATACTCTGTTATAGATAAACAAGACTAACTACTTGTTGGACACATGCCAAT	: : : :	237	
Db	137	LeuArgTyrGluIlePheLysThrGlySer	: : : :	152	
QY	238	TMTTGGAACTGAGTTCACTAGTAAGTACCAGCTGGTCTGCTCTCTTTTCTCGTAG	: : : :	297	
Db	153	PheThrSerProAsnGlyThrIleGluSerProGly	: : : :	167	
QY	298	ATATCTGGAACAATCTCCATGATATAGATATCCCTGCCATTTGTCATGGCTGGTGTG	: : : :	357	
Db	168	-----LysTyrProHisAsnLeu--AspCysThrPheThrIleLeuAlaLysPro	: : : :	183	
QY	358	CATCAGAGTAGTGTCAAACACCTGTGGGGGGCCAATCAGT	: : : :	399	
Db	184	LysMetGluIleLeuGlnPheLeuIlePheAspLeuGluHisAspProLeuGlnVal	: : : :	203	
QY	400	-----GTTGTAATTAGTAAGGTATCCC	: : : :	423	
Db	204	GlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHisValGly	: : : :	223	
QY	424	-----TATTATGAAGTCTTTGGCTAACACGTCACATCTGTGTGGGA	: : : :	468	
Db	224	ProLeuIleGlyLysTyrCysGlyThrLysThrProSerGluLeuArgSerSerThrGly	: : : :	243	
QY	469	CACATTATCTCAAGCTCTTTTACA	: : : :	492	
Db	244	IleLeuSerLeuThrPheHisThrAspMetAlaValAlaLysAspGlyPheSerAlaArg	: : : :	263	
QY	493	-----TTTAAGACAAGTGGATGTTATGGAACTGGGATG	: : : :	528	
Db	264	TyrTyrLeuValHisGlnGluProLeuGluAsnPheGlnCysAsnValProLeuGlyMet	: : : :	283	
QY	529	GAGTGTGTGTGATCGGGATCCCTCAATAACAGCATCATCTGTCTGGAGTGGACTGAC	: : : :	588	
Db	284	GluSerGlyArgIleAlaAsnGlnIleSerAlaSer	: : : :	297	
QY	589	CACACAGGGCAAGAAGTGTGAAACCCAAAAAGCCAGGCTGAAAAACCTGGACCG	: : : :	648	
Db	298	---ThrTyrSerAspGlyArgTyrThrProGlnSerArgLeuHisGlyAspAspAsn	: : : :	316	
QY	649	CCTTGGGCTGTTTGGCCTGATGAATACAGTGGTTACAAATAGATTTGNATGAAGAA	: : : :	708	
Db	317	GlyTrpThrProAsnLeuAspSerAsnLysGlyTyrLeuGlnValAspLeuArgPheLeu	: : : :	336	
QY	709	AAGAAATAACAGGCAATATACCACTGGA	: : : :	762	
Db	337	ThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGluThrGlnAsnGlyTyr	: : : :	356	
QY	763	TATGTGCTGCTACAGAATCCTGTACAGTATGATGGCAGAAATGGACTGTGTACAGA	: : : :	822	
Db	357	TyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAspTrpMetValTyrArg	: : : :	376	
QY	823	GAGCTGTGTGGACAGATAGAATATTTCAAGGAACAAGAATATACACCATGATGTG	: : : :	882	
Db	377	HisGly-----LysAsnHisLysValPheGlnAlaAsnAspAlaThrGluValVal	: : : :	394	
QY	883	CGTAACTACTTTTGGCCACCAATATTGCACGTTTTTATAGAGTGAATCCTACCCAATGG	: : : :	942	
Db	395	LeuAsnLysLeuHisAlaProLeuLeuThrArgPheValArgIleArgProGlnThrTrp	: : : :	414	
QY	943	CAGCAGAAAATTGGCATGAAATGGAGCTGTCTGGATGTCAAGTTTATTTCTCTAAAGGTGCT	: : : :	1002	

Db 736 laserGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlyGlyGluTrpLys 755
 QY 1843 GCTGGAGCCAGCTTACTGTCGCCAGAGCAATGTGTACCAAGTGCACAGAGCACA 1902
 Db 756 HisGlyArgIleLeuProSerTyrAsp---MetGluTyrGlnIle-----ValPhe 772
 QY 1903 CAAGAAGTATCAGGAGCAGGAAGGATGGGAA 1935
 Db 773 GlucyValIleGlyLysGlyArgSerGlyGlu 783
 RESULT 15
 NRPI_CHICK
 ID NRPI_CHICK STANDARD; PRT; 914 AA.
 AC P79795;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neuropilin-1 precursor (A5 protein).
 GN NRPI OR NRP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=White leghorn; TISSUE=Embryonic brain;
 RX MEDLINE=95324761; PubMed=7601310;
 RA Takagi S., Kasuya Y., Shimizu M., Matsuura T., Tsuboi M., Kawakami A.,
 RA Fujisawa H.;
 RT "Expression of a cell adhesion molecule, neuropilin, in the
 RT developing chick nervous system.";
 RL Dev. Biol. 170:207-222(1995).
 CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
 CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
 CC SIMILARITY). SEEMS TO HAVE CALCIUM-INDEPENDENT CELL ADHESION
 CC PROPERTIES.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: DEVELOPING NERVOUS SYSTEM; OPTIC TECTUM
 CC (LAYERS D AND E OF SGF), AMACRINE CELLS OF RETINA, NEURITES OF
 CC DORSAL ROOT GANGLIA. ALSO EXPRESSED IN NONNEURONAL CELLS, E.G.
 CC BLOOD VESSELS IN THE ENTIRE EMBRIO.
 CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 MAM-DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D45416; BAA08256.1; -;
 DR HSSP; P12259; 1CZT.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR000998; MAM_domain.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00629; MAM; 1.
 DR Pfam; PF00754; F5_F8_Type_C; 2.
 DR PRINTS; PR00020; MAMDOMAIN.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; SM00137; MAM; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00740; MAM_1; 1.

DR PROSITE; PS00060; MAM_2; 1.
 KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
 FT Cell adhesion.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 914 NEUROPILIN-1.
 FT DOMAIN 20 847 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 848 870 POTENTIAL.
 FT DOMAIN 871 914 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 25 139 CUB 1.
 FT DOMAIN 145 263 CUB 2.
 FT DOMAIN 273 422 F5/8 TYPE C 1.
 FT DOMAIN 429 581 F5/8 TYPE C 2.
 FT DOMAIN 636 801 MAM.
 FT DISULFID 25 52 PROBABLE.
 FT DISULFID 80 102 PROBABLE.
 FT DISULFID 145 171 PROBABLE.
 FT DISULFID 204 226 PROBABLE.
 FT DISULFID 273 422 BY SIMILARITY.
 FT DISULFID 429 581 BY SIMILARITY.
 SQ SEQUENCE 914 AA; 102480 MW; DD2EE6D6F0CBB68C CRC64;
 Alignment Scores:
 Pred. No.: 2.05e-13 Length: 914
 Score: 284.50 Matches: 135
 Percent Similarity: 34.35% Conservative: 79
 Best Local Similarity: 21.67% Mismatches: 190
 Query Match: 7.99% Indels: 219
 DB: 1 Gaps: 24
 US-10-060-830-1113 (1-1962) x NRPI_CHICK (1-914)
 QY 49 GACGCTGGAGCCAGCAAGCAATAGTCTGGTGGTGGCAATGAACCACTCA--- 105
 Db 217 AspValGlyProHisIleGlyArgTyrCysGly-----GlnAsnAsnProGlyArg 233
 QY 106 ATTGAATCAAAAGGCAATGAATCAGTCTGTGTCATGAGTGAATCCATGTTCTCGA 165
 Db 234 ValArgSerThrGlyLeuSerMetValPheTyrThrAspSerAlaIleAlaLys 253
 QY 166 CGCGGATTTTGGCTCATCTCTGTATAGATAAACAAGATCAATTAATTCTTTGGAC 225
 Db 254 GluGlyPheSerAlaAsnTyrSerVal----- 262
 QY 226 ACTGATCCCAATTTTGGAACTGAGTTCAGTAGTACTGCCAGCTGGTGTCTGCTT 285
 Db 263 ---SerGlnSerSerValSerGluAspPheGln-----CysMetGlu 275
 QY 286 CCTTTGCTGAGATATCTGGAACAATTCCTCATGATATAGATTCCTGCCCATTTGTCG 345
 Db 276 ProLeuGlyMetGluSerGlyGluIle-----HisSerAspGlnIleThr 290
 QY 346 ATGCTGGTGTGTCATGCGAGGAGTAGTGTCAACACCGTTGGGC----- 387
 Db 291 ValSerGlnTyrSerAlaIleTrpSerSerGluArgSerArgLeuAsnTyrProGlu 310
 QY 388 -----GGCAAAATCAGTGT----- 402
 Db 311 AsnGlyTrpThrProGlyGluAspSerValArgGluTrpIleGlnValAspLeuGlyLeu 330
 QY 403 -----GTAATTAGTAAGGTATT-----CCC 423
 Db 331 LeuArgPheValSerGlyIleGlyThrGlnGlyAlaIleSerLysGluThrLysLysGlu 350
 QY 424 TATTATGAAAGTCTCTTGGCTAAACAACGTCACATCT----- 459
 Db 351 TyrTrpLeuLysThrTyrArgValAspValSerSerAsnGlyGluAspTrpIleThrLeu 370
 QY 459 ----- 459
 Db 371 LysGluGlyAsnLysProValValPheGlnGlyAsnSerAsnProThrAspValValTyr 390
 QY 460 -----GTGGTGGGACAC 471
 III

Db 391 ArgProPheProLysProValLeuThrArgPheValArgIleLysProValSerTrpGlu 410
QY 472 TTATCTACAAAGTCTTTTACATTTTAAGACAAGTGA-----TGT 510
Db 411 AsnGlyValSerLeu---ArgPheGluValTyrglyCysLysIleThrAspTyProCys 429
QY 511 TATGGAACACTGGGATGAGTGTGTGTGATCGCGGATCCTCAATAACACAGCATCATCT 570
Db 430 SerGlyMetGluGlyMetValSerGlyLeuIleProAspSerGlnIleThrAlaSer--- 448
QY 571 GTGCTGGAGTGGACTGACACACAGGCGGCAAGAGACAGTGGAAACCCCAAAAAGCCAGG 630
Db 449 -----ThrGlnValAspArgAsnTrpIleProGluAsnAlaArg 461
QY 631 CTGAAAAACCTGGACCGCTGGCGTGTGCTTTTGGCCACTGATGAA-----TACCAG 681
Db 462 LeuIleThrSerArgSerGlyTrpAlaLeuProProThrHisProTyThrAsnGlu 481
QY 682 TGGTTTCAATAGATTGAATAGGAAAGAAATAACAGGCATTAATACCACTGGATCC 741
Db 482 TrpLeuGlnIleAspLeuGlyGluGlyIleValArgGlyIleIleValGlnGlyGly 501
QY 742 ACCATGGTGGACACAATTAATGTCTGTCTGCTCAGCAATCTGTACAGTGTATGGG 801
Db 502 LysHisArgGluAsnLysValPheMetLysLysPheLysIleGlyTySerAsnAsnGly 521
QY 802 CAGAAATGGACTGTGTACAGAGAGCCTGTGTGGACACAGATAAGATATTCAAGAAAC 861
Db 522 SerAspTrpLysMetIleMetAspSerSerLysLysLysIleLysThrPheGluGlyAsn 541
QY 862 AAGATTATCACACGAGTGTGGTAATACTTTTCCACCAATTAATGCACGTTTTATT 921
Db 542 ThrAsnTyAspThrProGluLeuArgThrPhe---GluProValSerThrArgIleIle 560
QY 922 AGAGTGAATCCTACCCAA---TGGCAGCAGAAATGCCATGAAATGGAGTGTCTCGGA 978
Db 561 ArgValTyProGluArgAlaThrHisAlaGlyLeuGlyLeuArgMetGluLeuGly 580
QY 979 TGTCAAGTTTATTCCTAAAGTGTCTCTCCAAACTTACTCAACCTCCACCTCTCGAAC 1038
Db 581 Cys----- 581
QY 1039 AGCAATGACCTCAAAACACTACAGCCCTCCAAATAATAGCCAAAGTGTGCCCAAAA 1098
Db 582 -----GluLeuGluAlaProThrAlaValProThrValSerGluGlyLysProValAsp 599
QY 1099 TTTACGCAACCACTACAACTCCAGTAGCAATGAATTTCTCAGACAGACAGAACAAACA 1158
Db 600 GluCysAspAspGlnAlaAsnCysHisSerGlyThrGlyThrThrValLeuAsn 619
QY 1159 ACTGCCAGTCTGATATCAGAAATACTACGGTAACCTCAAAATGTAACCAAGATGTAGCG 1218
Db 620 ThrGluLysProThrValIleAspAsnThrValGlnProGlu----- 633
QY 1219 CTGGCTGACGTTCTGTGCTGTGCTGGTGTGCTGCTCCTACTCTCTCATATTA 1278
Db 634 -----LeuProPro-----TyrAsnLeu 639
QY 1279 GTGTGTCTTGGCACTGGAGAAACAGAAAGAAAAAACTGAAGGCACCTATGACTTA--- 1335
Db 640 AsnCysGlyPheGlyTrpGlySerHisLysThrLeuCysGlnTrpGluHisAspAsnGln 659
QY 1336 -----CCTTACTGGGACCGGCGCAGGT 1356
Db 660 ValAspLeuLysTrpAlaIleLeuThrSerLysThrGlyProIleGlnAspHisThr--- 678
QY 1357 TGTGGAAGAGCAATGAAGCATTTCTCCGCAAAAGCAGTGGACCATGAGGAA----- 1410
Db 679 -----GlyAspGlyAsnPheIleTySerGlnAlaAspGluSerGlnLysGlyLys 695
QY 1411 -----ACCCGAGTTCGTATGACGACGACGCAAGTTAAT----- 1443
Db 696 ValAlaArgLeuLeuSerProIleIleTySerGlnAsnSerAlaHisCysMetThrPhe 715

QY 1444 -----CACCTGAGTCCAAAGAGAAGTCACCACAGTGTCTGACGAGTGTCTGACAGATAT 1497
Db 716 TrpTyThrHisMetSer----- 720
QY 1498 GCTCAGCCACTGGTAGGAGGAATTTGTTGTACACTTTCATCAAGATCTACCTTT---AAA 1554
Db 721 -----GlyProHisValGlyThrLeuLysIleLysLeuArgTyGlnLys 735
QY 1555 CCAGAAGAA 1563
Db 736 ProAspGlu 738

Search completed: January 21, 2003, 09:38:05
Job time : 37.5369 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 10:13:26 ; Search time 175.276 Seconds
(without alignments)
3854.498 Million cell updates/sec

Title: US-10-060-830-4
Perfect score: 300
Sequence: 1 aaacagaaagaaaaactg.....tatgcagacccctagctcctta 300

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	300	100.0	2046	21 AAZ51872	Human Factor 8 Hom
2	230	76.7	594	23 AAS63485	DNA encoding novel
3	183	61.0	640	21 AAZ80396	Human colon cancer
4	109	36.3	580	22 ABA60693	Human foetal liver
5	109	36.3	580	22 AAK08975	Human brain expres
6	109	36.3	580	22 AAK34866	Human bone marrow
7	109	36.3	580	22 AAI40582	Probe #9268 used t
8	106	35.3	467	22 ABA76819	Human foetal liver
9	106	35.3	467	22 AAK25439	Human brain expres

10	106	35.3	467	22	AAK51451	Human bone marrow
11	106	35.3	467	22	AAI57541	Probe #26227 used
12	58	19.3	1907	24	ABD35994	Human neuropilin-H
13	58	19.3	1907	24	ABK49567	Human cDNA encoding
14	58	19.3	1907	24	ABN59799	Novel human coding
15	58	19.3	2428	22	ABK34855	Human colon cancer
16	20	6.7	393	24	ABN66437	Streptococcus poly
17	20	6.7	440	21	AAF11554	Staphylococcus niger
18	20	6.7	11091	24	ABN92780	Staphylococcus epi
19	19	6.3	423	22	ABA48060	Human breast cell
20	19	6.3	423	22	ABA65943	Human foetal liver
21	19	6.3	423	22	ABA33035	Human foetal liver
22	19	6.3	423	22	AAK14364	Human brain expres
23	19	6.3	423	22	AAK40093	Human bone marrow
24	19	6.3	423	22	AAI20873	Probe #10806 for g
25	19	6.3	423	22	AAI46111	Probe #14797 used
26	19	6.3	423	22	AAI06583	Probe #6574 used t
27	19	6.3	423	22	ABSL4168	Human genome-deriv
28	19	6.3	1497	22	ABK87499	Human immune/haema
29	19	6.3	2122	23	AAS80960	DNA encoding novel
30	19	6.3	2521	23	AAS66068	DNA encoding novel
31	19	6.3	7105	22	AAI37450	Human musculoskele
32	19	6.3	13585	17	AAI11549	Tumour rejection a
33	18	6.0	30365	22	AAI20405	P. syringae pv. to
34	18	6.0	286	21	ABU53732	Arabidopsis thalia
35	18	6.0	370	24	ABL77512	Human ovarian canc
36	18	6.0	452	24	ABL93879	Arabidopsis thalia
37	18	6.0	581	22	AAI44194	Probe #12880 used
38	18	6.0	789	20	AAI13443	Enterococcus faeca
39	18	6.0	2708	22	AAH17390	Human cDNA sequenc
40	18	6.0	3232	21	ABU58202	Polyhydroxyalkanoa
41	18	6.0	4240	24	ABI99878	Mouse ischaemic co
42	18	6.0	4812	22	AAI04812	Human reproductive
43	18	6.0	4812	22	ABL97706	Human testicular a
44	18	6.0	13021	23	AAI04666	Human reproductive
45	18	6.0	13021	23	ABL97573	Human testicular a

ALIGNMENTS

RESULT 1
AAZ51872 standard; cDNA: 2046 BP.
AAZ51872;

04-JUL-2000 (first entry)

Human Factor 8 Homologue cDNA.

Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic; cerebroprotective; therapeutic; coagulation related disorder; haemophilia; stroke; screening; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 19..2028

Product= "Factor 8 Homologue"

WO200012532-A1.

09-MAR-2000.

20-AUG-1999; 99WO-US19047.

31-AUG-1998; 98US-0098521.

(ELIL) LILLY & CO ELI.

Rostock PRJ, Su W, Li XM;

```
XX WPI; 2000-256580/22.
DR P-PSDB; AAY70539.
XX
XX Factor 8 homolog polypeptides and nucleic acids encoding them for
PT treating coagulation related disorders such as hemophilia and stroke
XX
XX Claim 1; Page 61-64; 68pp; English.
XX
XX The present sequence is a cDNA encoding human Factor 8 homologue (F8H)
CC which is a coagulation cofactor. The protein is selectively expressed in
CC haematopoietic, heart and reproductive tissues. It has haemostatic and
CC cerebroprotective activities. The F8H contains a Factor 5/8 signature
CC and is useful as a therapeutic for treating coagulation related diseases
CC such as haemophilia and stroke. The nucleic acid is useful as
CC hybridisation probe and amplification primer for detecting deficiencies
CC in the level of F8H mRNA, for screening F8H gene mutations and for
CC monitoring regulation of gene expression. Fragments of the nucleic acid
CC are also useful as diagnostic probes and primers, and can be used in
CC screening methods such as those using DNA chips. The present sequence is
CC also useful as a target to screen therapeutically useful modulators
CC of the F8H.
XX
XX Sequence 2046 BP; 615 A; 457 C; 467 G; 507 T; 0 other;
SQ
Query Match 100.0%; Score 300; DB 21; Length 2046;
Best Local Similarity 100.0%; Pred. No. 3e-136;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAACGAAAGAAAACCTGAAGGACCTATGACTTACTTACTGGACCGGCGAGTTG 60
DB 1365 AAACGAAAGAAAACCTGAAGGACCTATGACTTACTTACTGGACCGGCGAGTTG 1424
QY 61 GTGGAAAGGAATGAAGCAGTTTCTCTGCAAAAGCAGTGGACCATGAGGAACCCCACT 120
DB 1425 GTGGAAAGGAATGAAGCAGTTTCTCTGCAAAAGCAGTGGACCATGAGGAACCCCACT 1484
QY 121 TCGCTATAGCAGCGAAGTTAATCAGTCTGAGTCCAGAGAGTCCACACAGTCTGCA 180
DB 1485 TCGCTATAGCAGCGAAGTTAATCAGTCTGAGTCCAGAGAGTCCACACAGTCTGCA 1544
QY 181 GCGTACTCTGAGAGTATGCTCAGCAGTCTGAGGAGGAAGTGTGTACACTTCATCA 240
DB 1545 GCGTACTCTGAGAGTATGCTCAGCAGTCTGAGGAGGAAGTGTGTACACTTCATCA 1604
QY 241 AAGATCTACCTTTAAACCAAGAGAAAGAAAGAGCAGGCTATGACAGACTAGATCCCTTA 300
DB 1605 AAGATCTACCTTTAAACCAAGAGAAAGAAAGAGCAGGCTATGACAGACTAGATCCCTTA 1664
RESULT 2
AA565485
ID AA565485 standard; cdna; 594 BP.
XX
XX AA565485;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #1289.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
(HYSE-) HYSEQ INC.
DR Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG01298.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID No 1289; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 594 BP; 175 A; 160 C; 139 G; 120 T; 0 other;
SQ
Query Match 76.7%; Score 230; DB 23; Length 594;
Best Local Similarity 100.0%; Pred. No. 3.8e-102;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 71 ATGAAGCAGTTTCTCTGCAAAAGCAGTGGACCATGAGGAACCCCACTGCTATAGC 130
DB 1 ATGAAGCAGTTTCTCTGCAAAAGCAGTGGACCATGAGGAACCCCACTGCTATAGC 60
QY 131 AGCAGCGAAGTTAATCAGTCTGAGTCCAGAGAGTCCACAGTGTGCGAGGCTGACTCT 190
DB 61 AGCAGCGAAGTTAATCAGTCTGAGTCCAGAGAGTCCACAGTGTGCGAGGCTGACTCT 120
QY 191 GCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGTTGTCACCTTCATCAAGATCTACC 250
DB 121 GCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGTTGTCACCTTCATCAAGATCTACC 180
QY 251 TTTAAACCAAGAAAGAAAGAAAGAGCAGGCTATGACAGACCTAGATCCCTTA 300
DB 181 TTTAAACCAAGAAAGAAAGAAAGAGCAGGCTATGACAGACCTAGATCCCTTA 230
RESULT 3
AAZ80396/c
ID AAZ80396 standard; cdna; 640 BP.
XX
XX AAZ80396;
XX
XX 07-APR-2000 (first entry)
XX
XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:480.
XX
XX Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
```

RESULT 4
AAI40582
ID AAI40582 standard; DNA; 580 BP.
XX AAI40582;
XX
XX 17-OCT-2001 (first entry)
XX probe #9268 used to measure gene expression in human placenta sample.
XX Gene: Glyceraldehyde 3-phosphate dehydrogenase (GAPDH).
XX Tissue: placenta.

Human placenta sample.
disorder; ss. placenta; antenatal diagnosis.

NO-NAF-2000
NO-NAF-2599-A1.

Location/Qualifiers
19. #tag = #
Product - factor #

Location/Date
19.1.2000

Tag-28 amplifiers
product a "f"

Factor 3 Homology

/note="Vector: pCMV-SPORT6"
21..2252
/codon_start=1
/product="Similar to endothelial and smooth muscle
cell-derived neuropilin-like protein"
/protein_id="AAH29558.1"
/db_xref="GI:20988615"
/translation="MPLFLLLVLLLLLEDAGAQQGGGHTVLGPESGTLTSINYP
QTPNSTVCWEIRVMGERVRKFGDFIEDSDSCHFNFLRIYNGIVSRTEIGKYC
GLGLQNHSTIESKNEITLLFMGSIHVSGRGLASVYDKDLITCLDTASNFLEPE
FSKYCPAGCLLPFAEISGTHPHGRSDSPCLMAGVHAGVYNTLGGQISVYISGPIY
YESSLANNTSVVGHLSLSTFTKTCGCTGLMESGVIADPQITASSVLEWTDHTGQ
ENSKPKKARKLKKPPPAFAFDEYQWLIQIDLNKKEKTIITGTSTWYHNYYVA
YRIYSDGQKWTYREPVEQDKIFQGNKYHQDYVNNFELPIIARFIRVNPOTWQO
KIAMKELLGCGIFKPPKLPKTPPPRNSNDLKNTPAPKIAKGRAPFTQPLQPR
SSNEFFAQTEQTASPDIRNTTTPNTVDLAALVPLVLMVLTLLILVCAWHM
RNRKKTEGTYDLPMYDRAGFYLMVSLACHRNEGWMKQFLPAKADVHEETPVYS
SSEVNHLSPREVTTLQADSAEYAOPLVGGIVGTLHORSTFKPEEGKEAGYADLDPN
SPGOEYHAYAEPLPTGPEYATPIINDMSGHTTTSVGQSTSTFKATGNOPPLVGT
INTLLSRDSCSSAAQYDTPKAGKPLPAPDELVYQVPOSTQEVSGAGRGECDFEK
EIL"
BASE COUNT 1137 a 807 c 785 g 1129 t
ORIGIN

Query Match 65.0%; Score 195; DB 9; Length 3858;
Best Local Similarity 99.6%; Pred. No. 1.6e-99;
Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 55 AGGTTGGTGAAGAAGTGAAGCAGTTCCTCTCTCAAAAGCAGTGGACCATGAGGAAAC 114
Db 1643 AGGTTGGTGAAGAAGTGAAGCAGTTCCTCTCTCAAAAGCAGTGGACCATGAGGAAAC 1702
QY 115 CCCAGTTCGTATAGCAGCAGCGAAGTTAATCACCTGAGTCCAGAGAGAGTCCACACACT 174
Db 1703 CCCAGTTCGTATAGCAGCAGCGAAGTTAATCACCTGAGTCCAGAGAGAGTCCACACACT 1762
QY 175 GCTGAGGCTGACTCTGCAGAGTATGCTCAGCAGTGGTAGGAGGAATTTGGTGTACACT 234
Db 1763 GCTGAGGCTGACTCTGCAGAGTATGCTCAGCAGTGGTAGGAGGAATTTGGTGTACACT 1822
QY 235 TCATCAAGATCTACCTTTAAACCAAGAGAAAGAGCAGGCTATCGACACCTAGA 294
Db 1823 TCATCAAGATCTACCTTTAAACCAAGAGAAAGAGCAGGCTATCGACACCTAGA
QY 295 TCCTTA 300
Db 1883 TCCTTA 1888

RESULT 4
AC013497
LOCUS
DEFINITION
AC013497
AC013497.4 GI:7329299
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
AC013497 192247 bp DNA linear HTG 26-MAR-2000
Homo sapiens chromosome 3 clone RP11-319J24 map 3, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC013497
AC013497.4 GI:7329299
HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 192247)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 3, clone RP11-319J24
2 (bases 1 to 192247)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baird,T., Binkley,N., Binkley,N., Binkley,N., Binkley,N., Binkley,N.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6692347.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1852
Center clone name: 319_J-24

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182870 bases at least Q40
Consensus quality: 187032 bases at least Q30
Consensus quality: 188925 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 191247; sum-of-contigs
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 472: contig of 472 bp in length
* 473 572: gap of 100 bp
* 573 3945: contig of 3373 bp in length
* 3946 4045: gap of 100 bp
* 4046 11056: contig of 7011 bp in length
* 11057 11156: gap of 100 bp
* 11157 17728: contig of 6572 bp in length
* 17729 17828: gap of 100 bp
* 17829 30556: contig of 12728 bp in length
* 30557 30656: gap of 100 bp
* 30657 49294: contig of 18638 bp in length
* 49295 49394: gap of 100 bp
* 49395 64509: contig of 15115 bp in length
* 64510 64609: gap of 100 bp
* 64610 83042: contig of 18433 bp in length
* 83043 83142: gap of 100 bp
* 83143 112373: contig of 29231 bp in length
* 112374 112473: gap of 100 bp
* 112474 145663: contig of 33190 bp in length
* 145664 145763: gap of 100 bp
* 145764 192247: contig of 46484 bp in length.

```

FEATURES
source
Location/Qualifiers
1. .192247

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-319J24"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature 1. .472
/notes="assembly_fragment"

```

```

clone_end:SP6
vector_side:right
573. .3945
/notes="assembly_fragment"
4046. .11056
/notes="assembly_fragment"
11157. .17728
/notes="assembly_fragment"
17829. .30556
/notes="assembly_fragment"
30657. .49294
/notes="assembly_fragment"
49395. .64509
/notes="assembly_fragment"
64610. .83042
/notes="assembly_fragment"
clone_end:T7
vector_side:left
83143. .112373
/notes="assembly_fragment"
112474. .145663
/notes="assembly_fragment"
145764. .192247
/notes="assembly_fragment"
BASE COUNT 58524 a 38157 c 36840 g 57721 t 1005 others
ORIGIN

Query Match 46.7%; Score 140; DB 2; Length 192247;
Best Local Similarity 100.0%; Pred. No. 2.8e-68;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 AGTTGGTGGAAAGCAATGAAGCAGTTCCTCTGCAAAAGCAGTGGACCATGAGGAAC 114
Db 164101 AGTTGGTGGAAAGCAATGAAGCAGTTCCTCTGCAAAAGCAGTGGACCATGAGGAAC 164160

Qy 115 CCCAGTTGCGCTATAGCAGCAGCGAAGTAACTACCTGAGTCCAAAGAGAGTCACACAGT 174
Db 164161 CCCAGTTGCGCTATAGCAGCAGCGAAGTAACTACCTGAGTCCAAAGAGAGTCACACAGT 164220

Qy 175 GCTGCGAGCTGACTCTGCAG 194
Db 164221 GCTGCGAGCTGACTCTGCAG 164240

RESULT 5
AC106728 193623 bp DNA linear PRI 25-MAY-2002
LOCUS
DEFINITION
BAC Library complete sequence.
ACCESSION
AC106728
VERSION
AC106728.6 GI:21206081
KEYWORDS
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 193623)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbaria, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X.,

```

JOURNAL
REFERENCE
AUTHORS

Unpublished

2 (bases 1 to 192247)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B., Brown, A., Castie, A., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Devar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Mar 26, 2000 this sequence version replaced gi:6692347.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L1852
Center clone name: 319_J_24

----- Summary Statistics

Sequencing vector: M13: M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 182870 bases at least Q40

Consensus quality: 187032 bases at least Q30

Consensus quality: 188925 bases at least Q20

Insert size: 194000; agarose-fp

Insert size: 191247; sum-of-contigs

Quality coverage: 5.0 in Q20 bases; agarose-fp

Quality coverage: 5.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 472: contig of 472 bp in length

473 572: gap of 100 bp

573 3945: contig of 3373 bp in length

3946 4045: gap of 100 bp

4046 11056: contig of 7011 bp in length

11057 11156: gap of 100 bp

11157 17728: contig of 6572 bp in length

17729 17828: gap of 100 bp

17829 30556: contig of 12728 bp in length

30557 30556: gap of 100 bp

30557 49294: contig of 18638 bp in length

49295 49394: gap of 100 bp

49395 64509: contig of 15115 bp in length

64510 64609: gap of 100 bp

64610 83042: contig of 18433 bp in length

83043 83142: gap of 100 bp

83143 112373: contig of 29231 bp in length

112374 112473: gap of 100 bp

112474 145663: contig of 33190 bp in length

145664 145763: gap of 100 bp

145764 192247: contig of 46484 bp in length.

Location/Qualifiers

FEATURES

RESULT 5
AC013497
LOCUS
DEFINITION
Homo sapiens chromosome 3 clone RP11-319J24 map 3, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC013497
AC013497
VERSION
AC013497.4 GI:7329299
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 192247)
AUTHORS
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE
Homo sapiens chromosome 3, clone RP11-319J24

0;

.64

)

.24

)

.284

)

.344

us-10-060-830-6.oli.rge

Wed Jan 22 14:58:46 2003

```
source
1. .192247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RPI1-319J24"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature
1. 472
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
573. .3945
/note="assembly_fragment"
4046. .11056
/note="assembly_fragment"
11157. .17728
/note="assembly_fragment"
17829. .30556
/note="assembly_fragment"
30657. .49294
/note="assembly_fragment"
49395. .64509
/note="assembly_fragment"
64610. .83042
/note="assembly_fragment"
clone_end:T7
vector_side:left"
83143. .112373
/note="assembly_fragment"
112474. .145663
/note="assembly_fragment"
145764. .192247
/note="assembly_fragment"

BASE COUNT 58524 a 38157 c 36840 g 57721 t 1005 others
ORIGIN
Query Match 100.0%; Score 275; DB 2; Length 192247;
Best Local Similarity 100.0%; Pred. No. 3.1e-153;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAACCTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAGGGTACGGGGACCAACCTCCCC 60
Db 165188 CAACCTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAGGGTACGGGGACCAACCTCCCC 165247

QY 61 CACTAGTGGGACATTACATACACTTCTCCAGGACTGACAGCTGCTCCTCAGCCAGG 120
Db 165248 CACTAGTGGGACATTACATACACTTCTCCAGGACTGACAGCTGCTCCTCAGCCAGG 165307

QY 121 CCACGTATGATACCCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCCAGCAATGGTGT 180
Db 165308 CCACGTATGATACCCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCCAGCAATGGTGT 165367

QY 181 ACCAGGTGCCACAGACACACAAAGATATCAGAGCAGAGGATGGGNAATGTGATG 240
Db 165368 ACCAGGTGCCACAGACACACAAAGATATCAGAGCAGAGGATGGGNAATGTGATG 165427

QY 241 TTTTAAAGAAATCTTTGAAGATGATGCTGCTTT 275
Db 165428 TTTTAAAGAAATCTTTGAAGATGATGCTGCTTT 165462
```

```
RESULT 6
AC106728 193623 bp DNA linear PRI 25-MAY-2002
LOCUS Homo sapiens 3 BAC RPI1-319J24 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC106728
VERSION AC106728.6 GI:21206081
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Primates; Hominidae; Homo.
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 21, 2003, 12:56:47 ; Search time 301.512 Seconds
(without alignments)
3222.856 Million cell updates/sec

Title: US-10-060-830-1115
Perfect score: 60
Sequence: 1 ctgctgctgcagacgctgg.....gtggacacactgtactagcc 60

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues
Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_esti.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_fod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length	Description
1	60	100.0	332	14 H80005
2	60	100.0	388	14 H85099
3	60	100.0	418	14 H95544
4	60	100.0	459	9 AA885289
5	60	100.0	470	9 AI767271
6	60	100.0	502	9 AI499115

7	60	100.0	504	9 AI565996
8	60	100.0	526	9 AI420312
9	60	100.0	566	12 BF434973
10	60	100.0	605	10 AW138866
11	60	100.0	693	12 BF732384
12	60	100.0	703	10 AW070902
13	60	100.0	908	12 BG178371
14	60	100.0	1061	14 BM905706
15	46	76.7	692	12 BF732861
16	35	58.3	424	9 AA460989
17	32	53.3	338	13 BI037778
18	32	53.3	543	12 BF691675
19	29	48.3	534	10 AV603143
20	22	36.7	427	9 AJ449261
21	22	36.7	530	9 AJ448166
22	22	36.7	667	9 AL586285
23	19	31.7	216	13 BI285467
24	18	30.0	454	13 BM430856
25	18	30.0	576	17 BH044206
26	18	30.0	662	13 BI232249
27	18	30.0	868	13 BG918584
28	18	30.0	980	9 AA908991
29	17	28.3	306	10 AV298498
30	17	28.3	329	13 BM086829
31	17	28.3	330	13 BM032259
32	17	28.3	346	9 AA318153
33	17	28.3	418	14 RI2756
34	17	28.3	462	10 BE252178
35	17	28.3	477	10 BB858068
36	17	28.3	481	9 AI170345
37	17	28.3	488	17 AZ231045
38	17	28.3	499	13 BM004612
39	17	28.3	500	13 BI373635
40	17	28.3	516	13 BM086841
41	17	28.3	527	13 BM583747
42	17	28.3	531	12 BG134859
43	17	28.3	543	12 BE71709
44	17	28.3	549	9 AL678252
45	17	28.3	561	17 AZ121328

ALIGNMENTS

RESULT 1
H80005
LOCUS
DEFINITION
Ys65d12.sl Soares retina N2b4HR Homo sapiens cDNA clone
IMAGE:219671.3, similar to SP:BMFL_HUMAN P13497 BONE MORPHOGENETIC
PROTEIN 1 PRECURSOR ;, mRNA sequence.
ACCESSION
H80005
VERSION
H80005.1 GI:1058094
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 332)
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilton RK

TITLE
JOURNAL
COMMENT
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 298
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Insert Length: 1092 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 298.

FEATURES

source

Location/Qualifiers
1..332
/organism="Homo sapiens"
/db_xref="GDB:3847664"
/db_xref="taxon:9606"
/clone="IMAGE:219671"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCCGCTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 85 a 67 c 89 g 88 t 3 others
ORIGIN
Query Match 100.0%; Score 60; DB 14; Length 332;
Best Local Similarity 100.0%; Pred. No. 2.7e-21;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGCTGCTCGAGGACGCTGGAGCCAGGATGATGGATGGACACACTGTACTAGGC 60
|||||
Db 25 CTGCTGCTCGAGGACGCTGGAGCCAGGATGATGGATGGACACACTGTACTAGGC 84
|||||

RESULT 2
H85099
LOCUS
DEFINITION
H85099 388 bp mRNA linear EST 14-NOV-1995
IMAGE:220160 3', mRNA sequence.

ACCESSION
H85099
VERSION
H85099.1 GI:1063842
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 388)
Hillier, L., Clark, N., Dubucque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 297
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 860 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 297.

TITLE
JOURNAL
COMMENT
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 297
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 860 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 297.

FEATURES

source

Location/Qualifiers
1..388
/organism="Homo sapiens"
/db_xref="GDB:3848153"
/db_xref="taxon:9606"
/clone="IMAGE:220160"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCCGCTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 91 a 93 c 100 g 102 t 2 others
ORIGIN
Query Match 100.0%; Score 60; DB 14; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.9e-21;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGCTGCTCGAGGACGCTGGAGCCAGGATGATGGATGGACACACTGTACTAGGC 60
|||||
Db 71 CTGCTGCTCGAGGACGCTGGAGCCAGGATGATGGATGGACACACTGTACTAGGC 130
|||||

RESULT 3
H99544
LOCUS
DEFINITION
H99544 418 bp mRNA linear EST 15-DEC-1995
IMAGE:263200 3', mRNA sequence.

ACCESSION
H99544
VERSION
H99544.1 GI:1124212
KEYWORDS
EST.
SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 418)
Hillier, L., Clark, N., Dubucque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 296
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1174 Std Error: 0.00
Seq primer: ml3 -40 forward
High quality sequence stop: 296.

FEATURES
source
Location/Qualifiers
1..418
/organism="Homo sapiens"
/db_xref="GDB:3872842"

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 13:00:06 ; Search time 37.2441 Seconds
(without alignments)
3292.167 Million cell updates/sec

Title: US-10-060-830-6
Perfect score: 275
Sequence: 1 caacttcagttggtcagccc.....tttgaagatgatgctgttt 275

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 50.0

Searched: 393868 seqs, 222934149 residues

Word size : 0

Total number of hits satisfying chosen parameters: 787736

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	275	100.0	580	10	US-09-864-761-12400
2	275	100.0	5657	9	US-09-974-298-96
3	257	93.5	467	10	US-09-864-761-30344
4	19	6.9	9417	10	US-09-764-855-171
5	18	6.5	9377	10	US-09-801-874-3
6	17	6.2	456	10	US-09-924-035A-677
7	17	6.2	450	10	US-09-880-107-537
8	17	6.2	577	10	US-09-864-761-12564
9	17	6.2	613	10	US-09-770-149-855
10	17	6.2	1230	10	US-09-764-847-1854
11	17	6.2	2529	10	US-09-764-847-1859
12	17	6.2	316	10	US-09-815-242-4020
13	17	6.2	4138	10	US-09-802-472B-5
14	17	6.2	5814	10	US-09-764-847-1860
15	17	6.2	26048	10	US-09-764-869-1856
16	17	6.2	26591	10	US-09-764-877-2878
17	16	5.8	189	10	US-09-864-761-31208
18	16	5.8	213	10	US-09-974-300-7540
19	16	5.8	223	9	US-09-954-531-677

20	16	5.8	223	9	US-09-954-531-1092
21	16	5.8	223	10	US-09-969-708-177
22	16	5.8	279	10	US-09-960-352-11414
23	16	5.8	356	10	US-09-604-287A-476
24	16	5.8	356	12	US-10-007-805-476
25	16	5.8	377	9	US-09-736-457-1096
26	16	5.8	377	9	US-09-902-941-1096
27	16	5.8	377	9	US-09-849-636-1096
28	16	5.8	456	10	US-09-864-761-16100
29	16	5.8	477	10	US-09-864-761-14675
30	16	5.8	479	10	US-09-864-761-1875
31	16	5.8	577	10	US-09-864-761-12010
32	16	5.8	634	10	US-09-764-855-15
33	16	5.8	659	10	US-09-879-536-656
34	16	5.8	717	10	US-09-925-300-1
35	16	5.8	725	10	US-09-770-149-129
36	16	5.8	823	10	US-09-925-299-145
37	16	5.8	1190	10	US-09-962-832-96
38	16	5.8	1299	9	US-10-001-887-31
39	16	5.8	1366	10	US-09-864-761-32593
40	16	5.8	1382	10	US-09-925-300-692
41	16	5.8	1656	10	US-09-925-301-125
42	16	5.8	1817	9	US-09-764-868-68
43	16	5.8	1876	10	US-09-604-287A-477
44	16	5.8	1876	12	US-10-007-805-477
45	16	5.8	2396	10	US-09-748-107-1

ALIGNMENTS

RESULT 1
US-09-864-761-12400
; Sequence 12400, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30

102e

;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 12400
;; LENGTH: 580
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC013497.4
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
US-09-864-761-12400

Query Match 100.0%; Score 275; DB 10; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.7e-134; Indels 0; Gaps 0;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAACCTCAGTGTGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAAACCAACCTCCCC 60
DB 281 CAACCTCAGTGTGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAAACCAACCTCCCC 340
QY 61 CACTAGTGGAACTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGG 120
DB 341 CACTAGTGGAACTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGG 400
QY 121 CCAGTATGATATACCCGAAAGCTGGGAAGCCAGGTCTACCTGCCAGAGCAATGCTGT 180
DB 401 CCAGTATGATATACCCGAAAGCTGGGAAGCCAGGTCTACCTGCCAGAGCAATGCTGT 460
QY 181 ACCAGTGCACAGACACACAGAGATATCAGGACGAGGAGGATGGGAATGTGTATG 240
DB 461 ACCAGTGCACAGACACACAGAGATATCAGGACGAGGAGGATGGGAATGTGTATG 520
QY 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275
DB 521 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 555

RESULT 2
US-09-974-298-96
;; Sequence 96, Application US/09974298
;; Patent No. US20020156263A1
;; GENERAL INFORMATION:
;; APPLICANT: Chen, Huel-Mei
;; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
;; FILE REFERENCE: PA-0037 P
;; CURRENT APPLICATION NUMBER: US/09/974,298
;; PRIOR FILING DATE: 2001-10-04
;; PRIOR APPLICATION NUMBER: 60/238,331
;; PRIOR FILING DATE: 2000-05-10
;; NUMBER OF SEQ ID NOS: 194
;; SOFTWARE: PERL Program
;; SEQ ID NO 96
;; LENGTH: 5657
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc.feature
;; OTHER INFORMATION: Incyte ID No. US20020156263A1 241227.17
;; NAME/KEY: unsure
;; LOCATION: 4516
;; OTHER INFORMATION: a, t, c, g, or other
US-09-974-298-96

Query Match 100.0%; Score 275; DB 9; Length 5657;
Best Local Similarity 100.0%; Pred. No. 3.1e-134; Indels 0; Gaps 0;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAACCTCAGTGTGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAAACCAACCTCCCC 60
DB 1931 CAACCTCAGTGTGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAAACCAACCTCCCC 1990
QY 61 CACTAGTGGAACTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGG 120
DB 1991 CACTAGTGGAACTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGG 2050
QY 121 CCAGTATGATATACCCGAAAGCTGGGAAGCCAGGTCTACCTGCCAGAGCAATGCTGT 180
DB 2051 CCAGTATGATATACCCGAAAGCTGGGAAGCCAGGTCTACCTGCCAGAGCAATGCTGT 2110
QY 181 ACCAGTGCACAGACACACAGAGATATCAGGACGAGGAGGATGGGAATGTGTATG 240
DB 2111 ACCAGTGCACAGACACACAGAGATATCAGGACGAGGAGGATGGGAATGTGTATG 2170
QY 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275
DB 2171 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 2205

RESULT 3
US-09-864-761-30344
;; Sequence 30344, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F
;; FILE REFERENCE: Aesomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; PRIOR FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 00:35:10 ; Search time 399.904 Seconds
(without alignments)
11048.706 Million cell updates/sec

Title: US-10-060-830-1113
Perfect score: 1962
Sequence: 1 atgcctctctctctctgct.....tttttaagaatactcttga 1962

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N: Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1895.8	96.6	2046	21	AAZ51872 Human Factor 8 Hom
2	1286.6	65.6	2428	22	AAH34855 Human colon cancer
3	1159.2	59.1	1907	24	AAH35994 Human neurophilin-H
4	1159.2	59.1	1907	24	ABR49567 Human cDNA encoding
5	1159.2	59.1	1907	24	ABN59799 Novel human coding
6	932.8	47.5	1388	24	ABK83640 Human cDNA differe
7	594	30.3	594	23	AAK565485 DNA encoding novel
8	544.6	27.8	640	21	AAZ80396 Human colon cancer
9	473.8	24.1	580	22	ABK60693 Human foetal liver

10	473.8	24.1	580	22	AAK08975	Human brain expres
11	473.8	24.1	580	22	AAK34866	Human bone marrow
12	473.8	24.1	580	22	AAI40582	Probe #9268 used t
13	467	23.8	467	22	ABA76819	Human foetal liver
14	467	23.8	467	22	AAK25439	Human brain expres
15	467	23.8	467	22	AAK51451	Human bone marrow
16	467	23.8	467	22	AAI57541	Probe #26227 used
17	204	10.4	1620	22	AAK00615	Novel human protei
18	204	10.4	1761	22	AAK00614	Novel human protei
19	204	10.4	1768	22	AAK00613	Novel human protei
20	204	10.4	3594	22	AAK00660	Human TANGO 229 CD
21	202.8	10.3	412	20	AAZ09008	Differentiation in
22	131.2	6.7	324	24	AAK35993	Human expressed se
23	131.2	6.7	324	24	ABK49566	Human Neurophilin-H
24	117.4	6.0	1871	21	AAA96736	Polynucleotide iso
25	91.6	4.7	6585	21	AAA60446	Murine factor V en
26	89.8	4.6	2209	22	AAH57553	Human brain cell s
27	89.8	4.6	4545	22	AAK26692	Human breast cance
28	89.8	4.6	4545	22	AAK26702	Human breast cance
29	89.8	4.6	4681	24	ABK86006	Human cDNA encoding
30	88.2	4.5	6893	20	AAZ32182	Human coagulation
31	86.6	4.4	6909	16	AAK03920	Human Factor-V wil
32	86.6	4.4	6909	21	AAA60448	Human factor V CDN
33	86.6	4.4	6909	21	AAZ39631	Human Vth aggregat
34	86.6	4.4	6909	22	AAK07368	Human DNA encoding
35	86.6	4.4	6909	24	ABN95777	Gene #2275 used to
36	86.6	4.4	6925	17	AAK33948	Human Factor V mut
37	86.6	4.4	6925	17	AAK33942	Human Factor V CDN
38	86.4	4.4	2303	18	AAK47338	Murine development
39	86.4	4.4	2303	20	AAK18506	Murine del-1 encod
40	80	4.1	479	22	AAH57345	Human brain specif
41	79.8	4.1	1780	18	AAK47343	Human developmenta
42	78.6	4.0	318	18	AAK47337	Human developmenta
43	78.6	4.0	318	20	AAK18509	Murine Del-1 probe
44	78.2	4.0	1780	20	AAK18508	Human del-1 encodi
45	76.8	3.9	1270	24	ABN96523	Gene #3021 used to

ALIGNMENTS

RESULT 1

AAZ51872

ID AAZ51872 standard; cDNA; 2046 BP.

AC AAZ51872;

DT 04-JUL-2000 (first entry)

DE Human Factor 8 Homologue cDNA.

XX Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW cerebroprotective; therapeutic; coagulation related disorder;
KW haemophilia; stroke; screening; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT 19..2028

FT /*tag= a

FT /product= "Factor 8 Homologue"

XX WO200012532-A1.

XX 09-MAR-2000.

XX 20-AUG-1999; 99WO-US19047.

XX 31-AUG-1998; 98US-0098521.

XX (ELIL) LILLY & CO ELI.

XX Rostock PRJ, Su W, Li XM;

Db 864 GTGGTGTATCGCGGATCTCTCAATTAACAGCATCA-CTGTGCTGGAGTGGAGTACACACACA 922
QY 595 GGGCAAGAGACAGTGTGGAAACCC-AAAAAGCCAGCGTGAATAAAACCTGGACCGCTTG 653
Db 923 GGGCAAGAGACAGTGTGGAAACCCAAAAAGCCAGCGTGAATAAAACCTGGACCGCTTG 982
QY 654 GGCTGCTTTTCCCACTGATGAATACCACTGGTGTACCAATAGATTGAATAGGAAAGAA 713
Db 983 GG-TGCTTTTCCCACTGATGAATACCACTGGTGTACCAATAGATTGAATAGGAAAGAA 1041
QY 714 AATACAGGCATTTATACCACTGATCCACCATGGTGGAGGACAAATTACTATGTGTCTGC 773
Db 1042 AATACAGGCATTTATACCACTGATCCACCATGGT-GAGCACAAATTACTATGTGTCTGC 1100
QY 774 CTACAGATCTCTGTACAGTGTATGGCAGCAAAATGGACTGTGTACAGAGAGCGTGT-G 832
Db 1101 CTACAGATCTCTGTACAGTGTATGGCAGCAAAATGGACTGTGTACAGAGAGCGTGTGC 1160
QY 833 TGGAGCAAGATAAGATATTTTCAAGSAAACAAA--GATTATCACCAGGAT-GTGGCTAATA 889
Db 1161 TGGAGCAAGATAAGATATTTTCAAGSAAACAAAAGAAATTTATCACCAGGATGTGTGCTAATA 1220
QY 890 ACTTTTTCGCCAATATTTATTCACGTTTTATAGAGTGAATCTTACCAATGGCAGCAGA 949
Db 1221 ACTTTTTCGCCAATATTTATTCACG-TTTATAGAGTGAATCTTACCAATGGCAGCAGA 1279
QY 950 AAATTGCCATGAATGAGTGTGCGATGTGCTGATTTTATCTTAAAGTGTGCTCTCAA 1009
Db 1280 AAATTGCCATGAATGAGTGTGCGATGTGCTGATTTTATCTTAAAGTGTGCTCTCAA 1339
QY 1010 AACTTACTCAACCTCCACTCTCGGAACAGCAATGACTTCAAAACACTACAGCCCTC 1069
Db 1340 AACTTACTCAACCTCCACTCTCGGAACAGCAATGACTTCAAAACACTACAGCCCTC 1398
QY 1070 CAAAAA---TAGCCAAAGTGTGCCCCAAATTTACCAACCACTACAACTGCGAGTA 1126
Db 1399 CAAAAAATTTAGCCAAAGTGTGCCCCAAATTTACCAACCACTACAACTGCGAGTA 1458
QY 1127 GCAATGAATTTCTGCAGACAGCAACAAACACTGCGAGTCTCTGATATCAGAAATACTA 1186
Db 1459 GCAATGAATTTCTGCAGACAGCAACAAACACTGCGAGTCTCTGATATCAGAAATACTA 1518
QY 1187 CGTAACTCCAAATGTAACCAAGATGTAGCGTGGCTGCGAGTCTTCTGCTGTGCTGG 1246
Db 1519 CGTAACTCCAAATGTAACCAAGATGTAGCGTGGCTGCGAGTCTTCTGCTGTGCTGG 1578
QY 1247 TCATGGTCTCTACTCTCTCATTTATATAGTGTGTGCTGGCACTGGGAAACAGAA 1306
Db 1579 TCATGGTCTCTACTCTCTCATTTATATAGTGTGTGCTGGCACTGGGAAACAGAA 1638
QY 1307 AGAAAAAATCTGAAGCACTATGACTTACCTTACTGGGACGGGCGAGT 1356
Db 1639 AGAAAAAATCTGAAGCACTATGACTTACCTTACTGGGACGGGCGAGT 1688

RESULT 4
ID ABK49567
XX ID ABK49567 standard; cDNA; 1907 BP.
AC ABK49567;
DT
XX 15-JUL-2002 (first entry)
DE Human cDNA encoding Neuropilin-Hy2.
XX Human; ss; gene; neuropilin-Hy2; chromosome 6q21; neuronal growth;
KW nerve regeneration; neurodegenerative disease; learning disorder;
KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
KW organ growth; nervous system lesion; cancer; cell proliferation;
KW cell differentiation; stem cell growth factor activity;
KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
KW platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;

XX reperfusion; food supplement; DNA microarray.
OS Homo sapiens.
XX Location/Qualifiers
PH 101..1258
FT /*tag= a
FT /product= "Neuropilin-Hy2"
FT /note= "This coding sequence is specifically claimed in
FT claim 1"
FT sig_peptide 101..160
FT /*tag= b
FT mat_peptide 161..1255
FT /*tag= c
FT /label= Mature_Neuropilin_Hy2
PN WO200222780-A2.
XX 21-MAR-2002.
XX 11-SEP-2001; 2001WO-US28590.
XX 11-SEP-2000; 2000US-0659671.
PR 06-SEP-2001; 2001US-0659671.
XX (TANG/) TANG T Y.
XX Tang TY;
XX WPI; 2002-351881/38.
DR P-PSDB; ABK79460.
XX New neuropilin-like polypeptides for diagnosing, preventing and
XX treating neurological conditions and disorders, cancers, and for
XX inducing angiogenesis and neovascularisation
XX Claim 1; Page 121-123; 144pp; English.
XX The invention relates to an isolated neuropilin-like polypeptide
XX including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
XX encoding the proteins and the coding regions of the cDNAs. Also included
XX is a nucleic acid array comprising the cDNAs attached to a surface used
XX for detecting full-matches or mismatches to the cDNAs. The genes
XX for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
XX The nucleic acid array is useful for detecting full-matches or mismatches
XX to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
XX in modulating neuronal growth, regenerative capacity, treating
XX neurodegenerative diseases, learning and memory disorders, diagnosing and
XX mapping genetic neuronal defects and degenerative diseases like
XX Alzheimer's disease, for inducing angiogenesis, and neovascularisation
XX and organ growth and development (e.g. the heart). The nervous system
XX disorders include lesions of central or peripheral nervous systems,
XX including traumatic lesions, ischaemic lesions, infectious lesions,
XX degenerative lesions, lesions associated with nutritional diseases or
XX disorders, neurological lesions, and lesions caused by toxic substances.
XX The neuropilin-like proteins and cDNAs are also useful as markers for
XX cancers. The neuropilin-like proteins are useful for regulating cell
XX proliferation, cell differentiation, stem cell growth factor activity,
XX for inducing proliferation of neural cells, regeneration of nerve and
XX brain tissue, for treatment of central and peripheral nervous system
XX diseases, and neuropathies, such as Parkinson's disease,
XX Huntington's disease, amyotrophic lateral sclerosis, to regulate
XX haematopoiesis and treat myeloid and lymphoid cell disorders, various
XX anaemias, and platelet disorders, such as thrombocytopaenia,
XX regeneration and treatment of lung or liver fibrosis, reperfusion
XX injury in various tissues and as a food supplement or molecular
XX weight marker. The cDNAs are useful in gene identification, genome
XX mapping, transgenics, as hybridisation probes, for primer design, for
XX gene chips and as a DNA antigen. The present sequence encodes
XX neuropilin-Hy2.
SQ Sequence 1907 BP; 548 A; 449 C; 406 G; 504 T; 0 other;

Db	1399	CAAAAAATTATAGCAAAAGGTCGTGCCCCAAAATTTACCCACCACCTACCACTCCAGTA	1458
Qy	1127	GCAATGAATTTCCCTGCACAGACAGAACAAACAACTGCCAGTCCTGATATCAGAAATACTA	1186
Db	1459	GCAATGAATTTCCCTGCACAGACAGAACAAACAACTGCCAGTCCTGATATCAGAAATACTA	1518
Qy	1187	CCGTAACCTCCAAATGTACCAAGATGTAGCGCTGGCTGCAGTCTTGTCCCTGTGCTGG	1246
Db	1519	CCGTAACCTCCAAATGTAAACCAAGATGTAGCGCTGGCTGCAGTCTTGTCCCTGTGCTGG	1578
Qy	1247	TCATGGTCTCCTACTACTCTCATTTCTCATATAGTGTGTGGCACGTGGAGAACAGAA	1306
Db	1579	TCATGGTCTCCTACTACTCTCATTTCTCATATAGTGTGTGGCACGTGGAGAACAGAA	1638
Qy	1307	AGAAAAAACTGAAGGCACCTATGACTTACCTTACTGGGACCGGGCAGGT	1356
Db	1639	AGAAAAAACTGAAGGCACCTATGACTTACCTTACTGGGACCGGGCAGGT	1688
RESULT 6			
ID	ABK83640		
XX	ABK83640 standard; cDNA; 1388 BP.		
AC	ABK83640;		
DT	14-AUG-2002 (first entry)		
XX	Human cDNA differentially expressed in granulocytic cells #211.		
DE			
XX	Human; ss; granulocytic cell; DNA chip; bacterial infection;		
KW	viral infection; parasitic infection; protozoal infection;		
KW	fungal infection; sterile inflammatory disease; psoriasis;		
KW	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;		
KW	cardiac reperfusion injury; renal reperfusion injury; ARDS;		
KW	adult respiratory distress syndrome; inflammatory bowel disease;		
KW	Crohn's disease; ulcerative colitis; periodontal disease;		
KW	granulocyte activation; chronic inflammation; allergy.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200228999-A2.		
XX			
PD	11-APR-2002.		
XX			
PF	03-OCT-2001; 2001WO-US30821.		
XX			
PR	03-OCT-2000; 2000US-237189P.		
XX			
PA	(GENE-) GENE LOGIC INC.		
XX			
PI	Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;		
XX	WPI; 2002-435328/46.		
DR			
XX			
PT	Detecting granulocyte activation by detecting differential expression		
PT	of genes associated with granulocyte activation, which serves as		
PT	diagnostic markers that is useful for monitoring disease states and		
PT	drug toxicity		
XX			
PS	Claim 1; SEQ ID No 211; 114pp; English.		
XX			
CC	The invention relates to detecting (M1) granulocyte (GC) activation		
CC	(GCA), by detecting the level of expression of gene(s) (Gs) identified by		
CC	DNA chip analysis as given in the specification, and comparing		
CC	the expression level to an expression level in an unactivated		
CC	GC, where differential expression of Gs is indicative of GCA.		
CC	Also included are modulating (M2) GA by contacting GC with an agent		
CC	that alters the expression of at least one gene in Gs; (2) screening (M3)		
CC	for an agent capable of modulating GCA or an inflammation (especially		
CC	chronic) in a tissue, an allergic response in a subject, exposure of a		
CC	subject to a pathogen or sterile inflammatory disease using the		
CC	gene expression profile; (3) detecting (M4) an inflammation (especially		
CC	chronic) in a tissue, an allergic response in a subject, exposure of a		

RESULT 8	
AAZ80396/C	
ID	AAZ80396 standard; cDNA; 640 BP.
XX	
XX	AAZ80396;
XX	
XX	07-APR-2000 (first entry)
XX	
XX	Human colon cancer cell line SW480 cDNA clone SEQ ID NO:480.
XX	
XX	Human; gene expression product; diagnosis: tumour; colon cancer;
KW	colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW	hyperplasia; ds.
XX	
XX	Homo sapiens.
OS	
PN	W09964576-A2.
XX	
PD	16-DEC-1999.
XX	
PF	09-JUN-1999; 99WO-IB01062.
XX	
PR	10-JUN-1998; 98US-0088801.
XX	
PA	(FARB) BAYER CORP.
XX	
PI	Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI	Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI	Schlegel R;
XX	
DR	WPI; 2000-087220/07.
XX	
PT	Novel nucleic acids, used to develop products for the diagnosis and
PT	treatment of disorders involving unwanted cell proliferation,
PT	particularly cancers, especially colon cancer
XX	
PS	Claim 15; Page 322; 469pp; English.
XX	
CC	AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC	the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC	cDNA clones can be used to generate antisense oligonucleotides which
CC	can be used for antisense therapy. Methods and products from the present
CC	invention can be used for identifying and/or classifying cancerous cells
CC	present in a human tumour, particularly in solid tumours, e.g.
CC	carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC	can be used for developing agents for the diagnosis and treatment of
CC	disorders involving unwanted cell proliferation, such as neoplasia,
CC	dysplasia or hyperplasia.
XX	
SQ	Sequence 640 BP; 148 A; 134 C; 160 G; 192 T; 6 other;
	Query Match 27.88; Score 544.6; DB 21; Length 640;
	Best Local Similarity 96.7%; Pred. No. 7e-161;
	Matches 616; Conservative 0; Mismatches 15; Indels 6; Gaps
Qy	893 TTTTGCCACCAATTATTGCAGCTTTTATTAGATGTAATCTCCCAATGGCAGCAGAAA 952
Db	632 TTTTGCCACCAATTATTGCAGC-TTTATTAGAGT-AATCCTTACCATGGCAGCAG-AA 576
Qy	953 TTGCCATGAAATGGAGCTGCTCGGATGTCAAGTATTATTCCTAAAGGTGCTCTCCAAAC 1012
Db	575 TTGCCATGAAAT-GAATGCTCGGATGTCAAGTATTATTCNTAAA-GTNGTCTCCAAAC 518
Qy	1013 TTACTCAACCTCCACCTCTCGGAAACAGCAATGACCTCAAAAACACTACAGCCCTCCAA 1072
Db	517 TTACTCAACCTCCACCTCTTNGNAACAGCAATGACCTCAAAAACACTACAGCCCTCC-A 459
Qy	1073 AAATAGCAAAAGTCTGGCCCCCAAAATTAGCAACCACTACAACTCGCAGTAGCAATG 1132
Db	458 AAATAGCAAAAGTCTGGCCCCCAAAATTAGCAACCACTACAACTTCGCGAGTAGCAATG 399
Qy	1133 AATTTCCTGCAACAGACAGAACAAACAACTGCCAGTCTCTGATATCAAGAAATACTACGGTAA 1192

```
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX Example 4; SEQ ID NO: 8966; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;
XX
XX Query Match 24.1%; Score 473.8; DB 22; Length 580;
XX Best Local Similarity 99.6%; Pred. No. 1.5e-138;
XX Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1486 TCTGCAGAGTATGCTCAGCCACTGCTAGGAGGAATTGTTGTACACTTCATCAAGATCT 1545
DB |||||||
DB 64 TCCACAGATGCTCAGCCACTGCTAGGAGGAATTGTTGTACACTTCATCAAGATCT 123
QY 1546 ACCTTTAAACCAGAGAAGGAAAGAACAGCGCTATGCAGACCTTAGTCTTACAACCTCA 1605
DB |||||||
DB 124 ACCTTTAAACCAGAGAAGGAAAGAACAGCGCTATGCAGACCTTAGTCTTACAACCTCA 183
QY 1606 CCAGGCGAGGAGTTTATCATGCTCTGCTGTAACCACTCCCAATTACGGGCGCTGAGTAT 1665
DB |||||||
DB 184 CCAGGCGAGGAGTTTATCATGCTCTGCTGTAACCACTCCCAATTACGGGCGCTGAGTAT 243
QY 1666 GCACCCCAATCATCATGGAGTATGCTAGGCGACCCCACTTCAGTTGGTTCAGCCCTCC 1725
DB |||||||
DB 244 GCACCCCAATCATCATGGAGTATGCTAGGCGACCCCACTTCAGTTGGTTCAGCCCTCC 303
QY 1726 ACATCCACTTTCAAGGCTACGGGGAACCACTCCCACTAGTGGGAACCTTACAATACA 1785
DB |||||||
DB 304 ACATCCACTTTCAAGGCTACGGGGAACCACTCCCACTAGTGGGAACCTTACAATACA 363
QY 1786 CTTCTCTCCAGGACTGACAGTCTCTCCTCAGCCAGGCGCCAGTATGATACCCCGAAGCT 1845
DB |||||||
DB 364 CTTCTCTCCAGGACTGACAGTCTCTCCTCAGCCAGGCGCCAGTATGATACCCCGAAGCT 423
QY 1846 GGGAGCGCAGGCTACCTGCGCCAGAGCAATTGGTGTACCAAGTGCACAGAGCACACAA 1905
DB |||||||
DB 424 GGGAGCGCAGGCTACCTGCGCCAGAGCAATTGGTGTACCAAGTGCACAGAGCACACAA 483
QY 1906 GAAGTATCAGGAGGAGGAGTGGGAATGTGATGTTTTTAAGAAATCCTTTGA 1962
DB |||||||
DB 484 GAAGTATCAGGAGGAGGAGTGGGAATGTGATGTTTTTAAGAAATCCTTTGA 540

RESULT 11
AAK34866
ID AAK34866 standard; DNA; 580 BP.
XX
XX AC AAK34866;
XX
XX DE 06-NOV-2001 (first entry)
XX Human bone marrow expressed single exon probe SEQ ID NO: 9423.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX Homo sapiens.
XX
```

```
PN WO200157276-A2.
XX
XX PD Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX PF WPI; 2001-483446/52.
XX
XX PR 30-JAN-2001; 2001WO-US006668.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX
XX PR 26-MAY-2000; 2000US-0207456.
XX
XX PR 30-JUN-2000; 2000US-0608408.
XX
XX PR 03-AUG-2000; 2000US-0632366.
XX
XX PR 21-SEP-2000; 2000US-0234687.
XX
XX PR 27-SEP-2000; 2000US-0236359.
XX
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488900/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX PS Example 4; SEQ ID NO: 9423; 658pp + Sequence Listing; English.
XX
XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;
XX
XX Query Match 24.1%; Score 473.8; DB 22; Length 580;
XX Best Local Similarity 99.6%; Pred. No. 1.5e-138;
XX Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1486 TCTGCAGAGTATGCTCAGCCACTGCTAGGAGGAATTGTTGTACACTTCATCAAGATCT 1545
DB |||||||
DB 64 TCCACAGATGCTCAGCCACTGCTAGGAGGAATTGTTGTACACTTCATCAAGATCT 123
QY 1546 ACCTTTAAACCAGAGAAGGAAAGAACAGCGCTATGCAGACCTTAGTCTTACAACCTCA 1605
DB |||||||
DB 124 ACCTTTAAACCAGAGAAGGAAAGAACAGCGCTATGCAGACCTTAGTCTTACAACCTCA 183
QY 1606 CCAGGCGAGGAGTTTATCATGCTCTGCTGTAACCACTCCCAATTACGGGCGCTGAGTAT 1665
DB |||||||
DB 184 CCAGGCGAGGAGTTTATCATGCTCTGCTGTAACCACTCCCAATTACGGGCGCTGAGTAT 243
QY 1666 GCACCCCAATCATCATGGAGTATGCTAGGCGACCCCACTTCAGTTGGTTCAGCCCTCC 1725
DB |||||||
QY 1726 ACATCCACTTTCAAGGCTACGGGGAACCACTCCCACTAGTGGGAACCTTACAATACA 1785
DB |||||||
DB 304 ACATCCACTTTCAAGGCTACGGGGAACCACTCCCACTAGTGGGAACCTTACAATACA 363
QY 1786 CTTCTCTCCAGGACTGACAGTCTCTCCTCAGCCAGGCGCCAGTATGATACCCCGAAGCT 1845
DB |||||||
DB 364 CTTCTCTCCAGGACTGACAGTCTCTCCTCAGCCAGGCGCCAGTATGATACCCCGAAGCT 423
QY 1846 GGGAGCGCAGGCTACCTGCGCCAGAGCAATTGGTGTACCAAGTGCACAGAGCACACAA 1905
DB |||||||
DB 424 GGGAGCGCAGGCTACCTGCGCCAGAGCAATTGGTGTACCAAGTGCACAGAGCACACAA 483
QY 1906 GAAGTATCAGGAGGAGGAGTGGGAATGTGATGTTTTTAAGAAATCCTTTGA 1962
DB |||||||
DB 484 GAAGTATCAGGAGGAGGAGTGGGAATGTGATGTTTTTAAGAAATCCTTTGA 540

RESULT 12
AAI40582
```

QY 1613 AGAAGTTTATCATGCTGTAACCACTCCCAATTTACGGGGCTGAGTATGCAACCC 1672
DB 121 AGAAGTTTATCATGCTGTAACCACTCCCAATTTACGGGGCTGAGTATGCAACCC 180
QY 1673 CAATCATATGAGCATGTCAGGGCACCACCACTTCAAGTTGTCAGCCCTCCACATCCA 1732
DB 181 CAATCATATGAGCATGTCAGGGCACCACCACTTCAAGTTGTCAGCCCTCCACATCCA 240
QY 1733 CTTTCAAGGCTACGGGAACCACTCCCACTAGTGGGAATCAATACACTTCTCT 1792
DB 241 CTTTCAAGGCTACGGGAACCACTCCCACTAGTGGGAATCAATACACTTCTCT 300
QY 1793 CCAGGACTGACAGTGTCTCTCAGCCAGCCAGCCAGTATGATACCCGAAAGCTGGGAAGC 1852
DB 301 CCAGGACTGACAGTGTCTCTCAGCCAGCCAGCCAGTATGATACCCGAAAGCTGGGAAGC 360
QY 1953 CAGTCTACCTGCCCCAGAGCAATGTTGTCAGAGTGCCACAGAGCACACAAGATAT 1912
DB 361 CAGTCTACCTGCCCCAGAGCAATGTTGTCAGAGTGCCACAGAGCACACAAGATAT 420
QY 1913 CAGGAGCAGGAAGGATGGGAATGTCATGTTTTTAAAGAAATCCTT 1959
DB 421 CAGGAGCAGGAAGGATGGGAATGTCATGTTTTTAAAGAAATCCTT 467
RESULT 14
AAK25439
ID AAK25439 standard; DNA; 467 BP.
XX AC AAK25439;
XX XX
DT 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe SEQ ID NO: 25430.
DE Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX Homo sapiens.
XX OS
PN WO200157275-A2.
XX 09-AUG-2001.
XX PF
XX 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX
XX WPI; 2001-483446/52.
XX XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX XX
XX Example 4; SEQ ID NO: 25430; 650pp + Sequence Listing; English.
XX PS
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.

XX Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;
SQ Query Match 23.8%; Score 467; DB 22; Length 467;
Best Local Similarity 100.0%; Pred. No. 1.8e-136;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1493 AGTATGCTCAGGCACATGCTAGGAGGAATTTGGTACACTTCAATCAAGATCTACCTTTA 1552
DB 1 AGTATGCTCAGGCACATGCTAGGAGGAATTTGGTACACTTCAATCAAGATCTACCTTTA 60
QY 1553 AACCCAGAAGAAAGAGAGGAGGCTATGACAGACCTAGATCCTTCAACTACCAAGGC 1612
DB 61 AACCCAGAAGAAAGAGAGGAGGCTATGACAGACCTAGATCCTTCAACTACCAAGGC 120
QY 1613 AGGAAGTTTATCATGCTGCTGTAACCACTCCCAATTTACGGGGCTGAGTATGCAACCC 1672
DB 121 AGGAAGTTTATCATGCTGCTGTAACCACTCCCAATTTACGGGGCTGAGTATGCAACCC 180
QY 1673 CAATCATATGAGCATGTCAGGGCACCACCACTTCAAGTTGTCAGCCCTCCACATCCA 1732
DB 181 CAATCATATGAGCATGTCAGGGCACCACCACTTCAAGTTGTCAGCCCTCCACATCCA 240
QY 1733 CTTTCAAGGCTACGGGAACCACTCCCACTAGTGGGAATCAATACACTTCTCT 1792
DB 241 CTTTCAAGGCTACGGGAACCACTCCCACTAGTGGGAATCAATACACTTCTCT 300
QY 1793 CCAGGACTGACAGTGTCTCTCAGCCAGCCAGCCAGTATGATACCCGAAAGCTGGGAAGC 1852
DB 301 CCAGGACTGACAGTGTCTCTCAGCCAGCCAGCCAGTATGATACCCGAAAGCTGGGAAGC 360
QY 1853 CAGGCTTACCTGCCCCAGAGCAATGTTGTCAGAGTGCCACAGAGCACACAAGATAT 1912
DB 361 CAGGCTTACCTGCCCCAGAGCAATGTTGTCAGAGTGCCACAGAGCACACAAGATAT 420
QY 1913 CAGGAGCAGGAAGGATGGGAATGTCATGTTTTTAAAGAAATCCTT 1959
DB 421 CAGGAGCAGGAAGGATGGGAATGTCATGTTTTTAAAGAAATCCTT 467
RESULT 15
AAK51451
ID AAK51451 standard; DNA; 467 BP.
XX AC AAK51451;
XX XX
DT 06-NOV-2001 (first entry)
XX Human bone marrow expressed single exon probe SEQ ID NO: 26008.
DE Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX Homo sapiens.
XX OS
PN WO200157276-A2.
XX XX
XX 09-AUG-2001.
XX PF
XX 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX

DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 26008; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;
SQ
Query Match 23.8%; Score 467; DB 22; Length 467;
Best Local Similarity 100.0%; Pred. No. 1.8e-136;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1493 AGTATGCTCAGCCACTGGTAGGAGGAATGTTGGTACACTTCATCAAGATCTACCTTTA 1552
Db 1 AGTATGCTCAGCCACTGGTAGGAGGAATGTTGGTACACTTCATCAAGATCTACCTTTA 60
QY 1553 AACCAAGAGAGGAAAGAGCAGGCTATGCAGACCTAGATCCTTCAACTCACCAGGGC 1612
Db 61 AACCAAGAGAGGAAAGAGCAGGCTATGCAGACCTAGATCCTTCAACTCACCAGGGC 120
QY 1613 AGGAAGTTTATCATGCTCTGCTGAACCACTCCCAATTTACGGGGCCTGAGTATGCAACCC 1672
Db 121 AGGAAGTTTATCATGCTCTGCTGAACCACTCCCAATTTACGGGGCCTGAGTATGCAACCC 180
QY 1673 CAATCATCATGGACATGTCAGGGCACCACCAACTTCAGTTGGTTCAGCCCTCCACATCCA 1732
Db 181 CAATCATCATGGACATGTCAGGGCACCACCAACTTCAGTTGGTTCAGCCCTCCACATCCA 240
QY 1733 CTTTCAAGGCTACGGGGAACCAACTCCCCACTAGTGGGAACCTACAATACACTTCTCT 1792
Db 241 CTTTCAAGGCTACGGGGAACCAACTCCCCACTAGTGGGAACCTACAATACACTTCTCT 300
QY 1793 CCAGGACTGACAGCTGCTCCTCAGCCAGCCAGCCAGCTATGATACCCGAAAGCTGGGAAGC 1852
Db 301 CCAGGACTGACAGCTGCTCCTCAGCCAGCCAGCCAGCTATGATACCCGAAAGCTGGGAAGC 360
QY 1853 CAGGTCTACCTGCCCCAGACGAATTTGGTGTACAGGTGCCAGAGCACACAAGATAT 1912
Db 361 CAGGTCTACCTGCCCCAGACGAATTTGGTGTACAGGTGCCAGAGCACACAAGATAT 420
QY 1913 CAGGAGCAGGAGGATGGGGAATGTGATGTTTTTAAAGAAATCCTT 1959
Db 421 CAGGAGCAGGAGGATGGGGAATGTGATGTTTTTAAAGAAATCCTT 467

Search completed: January 21, 2003, 02:15:55
Job time : 412.904 secs

subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection, parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at www.vivo.int/pub/published_pct_sequences.

[illegible]

QY	56	GAGCCGAGCAAGCAAAATAC	TGTGGTCTGGGGTTGCAAAATGAACCAATCAATGAATCAAA	115
DB	217	GAACTGAAATAGCGCAAAATAC	TGTGGTCTGGGGTTGCAAAATCAACCAATCAAAATGAATCAAA	276
QY	116	AAGCAATGAAATACACATGCTGT	TTCATGAGTGAATCCATGTTCTCGACGCGGATTTT	175
DB	277	AAGCAATGAATACACATGCTGT	TTCATGAGTGAATCCATGTTCTCGACGCGGATTTT	336
QY	176	TGGCCTCATCTCTGTATTAGATAAAC	-AAGATCTAATTACTTGTTTGGACACATGCATCC	234
DB	337	TGGCCTCATCTCTGTATTAGATAAAC	ATGAATAAGATCTAATTACTTGTTTGGACACATGCATCC	395
QY	235	AA-TTTTTTTGGAACTCGAGTTCA	GTAAGTACTG-CCGACGTGGTGTCTGCTTCCTTTTG	292
DB	397	AATTTTTTTGGAACTCGAGTTCA	GTAAGTACTGCCCCAGCTGCTGTCTTCCTTTTG	456
QY	293	CTGAGATATCTGGAAACAATTCCT	CAATGGAATAGAGATTCCTGGCAATGTCATGGCTG	352
DB	457	CTGAGATATCTGGAAACAATTCCT	CAATGGAATAGAGATTCCTGGCAATGTCATGGCTG	516
QY	353	GNGTGCATGCAGGAGTAGTGTCA	AAACAGTTGGCGGCCAAATCACTGTTGTAATTAGTA	412
DB	517	GNGTGCATGCAGGAGTAGTGTCA	AAACAGTTGGCGGCCAAATCACTGTTGTAATTAGTA	576
QY	413	AAAGTATTCCCTATTATGAAAGT	TTTGGCTTAACAACGTCACATCTGTGGTGGACACT	472
DB	577	AAAGTATTCCCTATTATGAAAGT	TTTGGCTTAACAACGTCACATCTGTGGTGGACACT	636
QY	473	T--ATCTACAAGTC-TTTTTTACA	NTTAAGACAAGTGGATGTTATGAAACACTGGGGATGG	529
DB	637	TGATACTACAAGTC-TTTTTTACA	NTTAAGACAAGTGGATGTTATGAAACACTGGGGATGG	696
QY	530	AGTCTGGT-GTGATCGGGATCCTC	CAAAATACAGCATCATCTGTCTGGAGTGGACTGAC	588
DB	697	AGTCTGGTGGTGTATCGCGGATC	CTCAATAACAGCATCA-CTGTGTGGAGTGGACTGAC	755
QY	589	CACACAGGCGAGAGAACAGTTGG	AAACCC-AAAAAGCCAGGCTGAAAAAACCTCGACC	647
DB	756	CACACAGGCGAGAGAACAGTTGG	AAACCCAAAAAGCCAGGCTGAAAAAACCTCGACC	815
QY	648	GCCTTGGGCTGCTTTTGGCCACT	GATGAATACCAGTGGTTACAAATAGATTTGAATAGA	707

Query Match	59.1%	Score 1159.2	DB 24	Length 1907
Best Local Similarity	98.3%	Pred. No. 0		
Matches 1288	Conservative	0	Mismatches	8
			Indels	14
			Gaps	11
QY	56	GAGCCGACGAGGCAAAATAC	TGCGTCTGGGTTGCAATGCAATCAATCAATGAATCAA	115
DB	384	GAACTGAATAGGCAAAATAC	TGCGTCTGGGTTGCAATGCAATCAATCAATGAATCAA	443
QY	116	AAGGCAATGAATACACATTC	GTGCTCATGTAGTGGAAATCCATGTTCTGGACGCGGATTTT	175
DB	444	AAGGCAATGAATACACATTC	GTGCTCATGTAGTGGAAATCCATGTTCTGGACGCGGATTTT	503
QY	176	TGGCCTCATACTCTGTTTAT	AGATAAACAAGATCTAATTTACITGTTTGACACTGCATCCA	235
DB	504	TGGCCTCATACTCTGTTTAT	AGATAAACAAGATCTAATTTACITGTTTGACACTGCATCCA	563
QY	236	ATTTTTCGGAACCTGAGTTC	AGTACTGCCAGCTGGTGTCTGCTCTCTTTTGTCTG	295
DB	564	ATTTTTCGGAACCTGAGTTC	AGTACTGCCAGCTGGTGTCTGCTCTCTTTTGTCTG	623
QY	296	AGATATCTGGAAATTCCTC	TATGGATATAGAGATTCCTCGCAATGTCATGGCTGGTG	355
DB	624	AGATATCTGGAAATTCCTC	TATGGATATAGAGATTCCTCGCAATGTCATGGCTGGTG	683
QY	336	TGCATGCGAGGATAGTGT	CAACACGTTGGGGCGGCCAAATCAGTGTGTAATAGTAAAG	415
DB	684	TGCATGCGAGGATAGTGT	CAACACGTTGGGGCGGCCAAATCAGTGTGTAATAGTAAAG	743
QY	416	GTATTCCTTATTTATGAA	AGTCTTTGGCTTAAACGCTCACATCTGTGGTGGGACACTTAT	475
DB	744	GTATTCCTTATTTATGAA	AGTCTTTGGCTTAAACGCTCACATCTGTGGTGGGACACTTAT	803
QY	476	CTACAAGTCTTTTACATTT	TAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTG	535
DB	804	CTACAAGTCTTTTACATTT	TAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTG	863
QY	536	GT-GTATCGCGGATCCTCA	ATAACAGCATCATCTGTGCTGAGTGGACTGACCACACA	594
DB	864	GTGTTGATCGCGATCCTCA	ATAACAGCATCA-CTGTGCTGAGTGGACTGACCACACA	922
QY	595	GGGCAAGAGAACAGTTGG	AAACCC-AAAAAGCCAGGCTGAAAAAACCCTGGACCGCCTG	653
DB	923	GGGCAAGAGAACAGTTGG	AAACCCAAAAAGCCAGGCTGAAAAAACCCTGGACCGCCTG	982
QY	654	GGCTGCTTTTGCACCTGAT	GAATACCAGTGGTTTACAAATAGATTTGAATTAAGGAAAAAGAA	713
DB	983	GG-TGCTTTTGCACCTGAT	GAATACCAGTGGTTTACAAATAGATTTGAATTAAGGAAAAAGAA	1041
QY	714	AATAACAGGCAATTAATAC	CACTGGATCCCATGTTGGAGACAAATTTACTATGTGTCTGC	773
DB	1042	AATAACAGGCAATTAATAC	CACTGGATCCCATGTTGGAGACAAATTTACTATGTGTCTGC	1100
QY	774	CTACAGAACTCTGACAGT	GCATGATGGCAGAAATGGACTGTGTACAGAGAGCCCTGGT-G	832
DB	1101	CTACAGAACTCTGACAGT	GCATGATGGCAGAAATGGACTGTGTACAGAGAGCCCTGGTGC	1160
QY	833	TGGAGCAAGATTAAGATAT	TTTCAAGAAACAAA--GATTATCACCAGGAT-GTGGCGTAATA	889
DB	1161	TGGAGCAAGATTAAGATAT	TTTCAAGAAACAAAATTTATCACCAGGATGGTGGCGTAATA	1220
QY	890	ACTTTTGGCCACCAATTA	TGACAGTTTATTAGAGTGAATCTTACCCAAATGGGACGACAGA	949
DB	1221	ACTTTTGGCCACCAATTA	TGACAG-TTTATTAGAGTGAATCTTACCCAAATGGGACGACAGA	1279
QY	950	AAATGGCATGAAAATGG	AGCTGCTCGGATCTGATTTATCCCTAAAGGTGCTCTCCCAA	1009
DB	1280	AAATGGCATGAAAATGG	AGCTGCTCGGATCTGATTTATCCCTAAAGGTGCTCTCCCAA	1339
QY	1010	AACTTACTCAACCTCCAC	CTCCTCGGAACAGCAATGACCTCAAAAACACTACAGCCCTC	1069
DB	1340	AACTTACTCAACCTCCAC	CTCCTCGGAACAGCAATGACCTC-AAAACACTACAGCCCTC	1398
QY	1070	CAAAAA---TAGCCAAAG	GGTGTGCCCCCAAAATTTACGCAACCCACTACAACTGGAGTA	1126

Db	1399	CAAAATTTAGCCAAAGGTCGTGCCCAAAATTTACGCAACCCACTACAACTCGCAGTA	1458
Qy	1127	GCAATGAATTTCTGTCACACAGACAGACAACAACTGCGAGTCTGTATATCAGAAATACTA	1186
Db	1459	GCAATGAATTTCTGTCACACAGACAGACAACAACTGCGAGTCTGTATATCAGAAATACTA	1518
Qy	1187	CCGTAACTCCAAATGTAAACCAAGATGTAGCGCTGGCTGCAGTTCCTTGCCCTGTGCTGG	1246
Db	1519	CCGTAACTCCAAATGTAAACCAAGATGTAGCGCTGGCTGCAGTTCCTTGCCCTGTGCTGG	1578
Qy	1247	TCATGGCTCCTCACTACTCTCTCATTTCTATATAGTGTGTGCTTGGCACTGGAGAAACAGAA	1306
Db	1579	TCATGGCTCCTCACTACTCTCTCATTTCTATATAGTGTGTGCTTGGCACTGGAGAAACAGAA	1638
Qy	1307	AGAAAAAATGAAGGCACCTATCACTTACTTACTGGGACCGGCAGGT	1356
Db	1639	AGAAAAAATGAAGGCACCTATCACTTACTTACTGGGACCGGCAGGT	1688
RESULT 5			
ABN	59799	standard; cDNA; 1907 BP.	
XX	AC	ABN59799;	
XX	DT	28-JUN-2002 (first entry)	
XX	DE	Novel human coding sequence SEQ ID NO: 210.	
XX	KW	Human; antianaemic; vulnervary; antiinflammatory; immunomodulator;	
KW	KW	antinfertility; cerebroprotective; cytotatic; rheumatic; gene therapy	
KW	KW	neuroprotective; antiparkinsonian; protein therapy; EST;	
KW	KW	expressed sequence tag; gene; ss.	
XX	OS	Homo sapiens.	
XX	PN	WO200222660-A2.	
XX	PD	21-MAR-2002.	
XX	PF	10-SEP-2001; 2001WO-US26015.	
XX	PR	11-SEP-2000; 2000US-0659671.	
XX	PA	{HYSE-} HYSEQ INC.	
XX	PI	Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;	
PI	PI	Xue AJ, Yang Y, Wehrman T, Drmanac RT;	
XX	XX	WPI; 2002-292408/33.	
DR	DR	P-PSDB; ABB97386.	
XX	PT	An isolated polynucleotide for treating diseases associated with its	
XX	PS	encoded polypeptide such as cancer and multiple sclerosis -	
XX	PS	Claim 1; SEQ ID NO 210; 509pp; English.	
XX	CC	The present invention provides the protein and coding sequences of 444	
CC	CC	novel human proteins. These were isolated from expressed sequences tags	
CC	CC	(ESTs). They can be used to stimulate cell growth, to regulate	
CC	CC	haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth	
CC	CC	e.g. in burn treatment, to regulate the immune system e.g. to treat	
CC	CC	multiple sclerosis, to regulate activin or inhibin e.g. to treat	
CC	CC	infertility, to regulate haemostasis or thrombolysis e.g. to treat	
CC	CC	stroke and cancer, to screen for drugs, to treat inflammatory condition	
CC	CC	e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.	
CC	CC	Parkinson's disease. The present sequence is a coding sequence of the	
XX	XX	invention.	
XX	SQ	Sequence 1907 BP; 548 A; 449 C; 406 G; 504 T; 0 other:	

Query Match 59.1%; Score 1159.2; DB 24; Length 1907;

XX WPI; 2000-256580/22.
DR p-PSDB; AAV70539.
XX
PT Factor 8 homolog polypeptides and nucleic acids encoding them for
PT treating coagulation related disorders such as hemophilia and stroke
XX
XX
PS Claim 1; Page 61-64; 68pp; English.
PS
CC The present sequence is a cDNA encoding human Factor 8 homologue (F8H)
CC which is a coagulation cofactor. The protein is selectively expressed in
CC haematopoietic, heart and reproductive tissues. It has haemostatic and
CC cerebroprotective activities. The F8H contains a Factor 5/8 signature
CC and is useful as a therapeutic for treating coagulation related diseases
CC such as haemophilia and stroke. The nucleic acid is useful as
CC hybridisation probe and amplification primer for detecting deficiencies
CC in the level of F8H mRNA, for screening F8H gene mutations and for
CC monitoring regulation of gene expression. Fragments of the nucleic acid
CC are also useful as diagnostic probes and primers, and can be used in
CC screening methods such as those using DNA chips. The present sequence is
CC also useful as a target to screen therapeutically useful modulators
CC of the F8H.
XX
SQ Sequence 2046 BP; 615 A; 457 C; 467 G; 507 T; 0 other;

Query Match 96.6%; Score 1895.8; DB 21; Length 2046;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1900; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 56 GAGCCAGCAAGCAATGCTGGTCTGGGTGCAAAATGAACCATTCATTAATCAAA 115
DB 122 GAATCAAAATAGGCAAAATGCTGGTCTGGGTGCAAAATGAACCATTCATTAATCAAA 181
QY 116 AAGCAATGAATCAGATGCTGTTTCATGAGTGAATCCATGTTTCTGGACGCGGATTTT 175
DB 182 AAGCAATGAATCAGATGCTGTTTCATGAGTGAATCCATGTTTCTGGACGCGGATTTT 241
QY 176 TGGCTCATGCTGTTATAGATAAACAAGATCTAATTAATTTGTTGGACACATGCTATCA 235
DB 242 TGGCTCATGCTGTTATAGATAAACAAGATCTAATTAATTTGTTGGACACATGCTATCA 301
QY 236 ATTTTGGAACTGAGTTGCTAGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 295
DB 302 ATTTTGGAACTGAGTTGCTAGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361
QY 296 AGATATCTGGAACAATTCCTCATGATATAGATATGCTGCCATATGTCATGCTGCTGCTG 355
DB 362 AGATATCTGGAACAATTCCTCATGATATAGATATGCTGCCATATGTCATGCTGCTGCTG 421
QY 356 TGCATGCAAGGAGTAGTGTCAACACAGTTGGCGGCCAAATCAGTGTGTTGTAATTAGTAAAG 415
DB 422 TGCATGCAAGGAGTAGTGTCAACACAGTTGGCGGCCAAATCAGTGTGTTGTAATTAGTAAAG 481
QY 416 GTATTCCTATTATGAAAGTCTTTTGGCTAAACAGTGCACATCTGTTGGTGGACACTTAT 475
DB 482 GTATTCCTATTATGAAAGTCTTTTGGCTAAACAGTGCACATCTGTTGGTGGACACTTAT 541
QY 476 CTACAAGTCTTTTACATTTTAAGACAATGATGTTATGGAACACTGGGGATGAGTCTG 535
DB 542 CTACAAGTCTTTTACATTTTAAGACAATGATGTTATGGAACACTGGGGATGAGTCTG 601
QY 536 GTGTGATCGCGGATCTCAAAATACAGCATCTGCTGCTGAGTGGACTGACACACAG 595
DB 602 GTGTGATCGCGGATCTCAAAATACAGCATCTGCTGCTGAGTGGACTGACACACAG 661
QY 596 GGCAGAGAACAGTTGGAAACCCAAAAGCCAGGCTGAAACAACTGGACCGCTTGGG 655
DB 662 GGCAGAGAACAGTTGGAAACCCAAAAGCCAGGCTGAAACAACTGGACCGCTTGGG 721
QY 656 CTGCTTTTGGCCACTGATGAATACCAAGTGGTTACAAATAGATTTGAATAGGAAAAAGAAA 715
DB 722 CTGCTTTTGGCCACTGATGAATACCAAGTGGTTACAAATAGATTTGAATAGGAAAAAGAAA 781

QY 716 TAACAGGCAATTATAACCACTGGATCCACCATGGTGGAGCACAAATTAATGTGTCTGCT 775
DB 782 TAACAGGCAATTATAACCACTGGATCCACCATGGTGGAGCACAAATTAATGTGTCTGCT 841
QY 776 ACAGAACTCTGTACAGTGTATGGCGAGAAATGGACTGTGTACAGAGACCTGTGTGG 835
DB 842 ACAGAACTCTGTACAGTGTATGGCGAGAAATGGACTGTGTACAGAGACCTGTGTGG 901
QY 836 ACAGAACTCTGTACAGTGTATGGCGAGAAATGGACTGTGTACAGAGACCTGTGTGG 895
DB 902 ACAGAACTCTGTACAGTGTATGGCGAGAAATGGACTGTGTACAGAGACCTGTGTGG 961
QY 896 TGCCACCAATTTGTCACCTGTTTATAGAGTGAATCTTACCCTACCAATGGCAGCAAAATG 955
DB 962 TGCCACCAATTTGTCACCTGTTTATAGAGTGAATCTTACCCTACCAATGGCAGCAAAATG 1021
QY 956 CATGAAATGGAGCTGCTGGATGTCAGTTTATTCCTAAAGTCTGCTTCCAAAATTTA 1015
DB 1022 CATGAAATGGAGCTGCTGGATGTCAGTTTATTCCTAAAGTCTGCTTCCAAAATTTA 1081
QY 1016 CTCAACTCCACTCTCTCGGAACAGCAATGACCTCAAAACACTACAGCCCTCCAAA 1075
DB 1082 CTCAACTCCACTCTCTCGGAACAGCAATGACCTCAAAACACTACAGCCCTCCAAA 1141
QY 1076 TAGCCAAAGTCTGCTGCCCAAAATTTACGCAACACTTACAACCTCGCAGTAGCAATGAAT 1135
DB 1142 TAGCCAAAGTCTGCTGCCCAAAATTTACGCAACACTTACAACCTCGCAGTAGCAATGAAT 1201
QY 1136 TTCTCTGCACAGACAGAACAACTGCCAGTCTCTGATATCAGAAATTAATTAACCTTA 1195
DB 1202 TTCTCTGCACAGACAGAACAACTGCCAGTCTCTGATATCAGAAATTAATTAACCTTA 1261
QY 1196 CAAATGTAACCAAGATGTAGCGCTGGCTGCGACTTCTTCTGCTGCTGCTGCTGCTG 1255
DB 1262 CAAATGTAACCAAGATGTAGCGCTGGCTGCGACTTCTTCTGCTGCTGCTGCTGCTG 1321
QY 1256 TCACCTCTCTCTCTCATATTAATAGTGTGCTTGGACACTGGAGAAACAGAAAGAAA 1315
DB 1322 TCACCTCTCTCTCTCATATTAATAGTGTGCTTGGACACTGGAGAAACAGAAAGAAA 1381
QY 1316 CTGAAGGCACTATGACTTACCTTACTTGGACCGGGAGGTTGGTGGAAAGGATGAAGC 1375
DB 1382 CTGAAGGCACTATGACTTACCTTACTTGGACCGGGAGGTTGGTGGAAAGGATGAAGC 1441
QY 1376 AGTTTCTCTGCAAAAGCAGTGGACCATCAGGAAACCCAGTTCGCTATAGCAGCAGCG 1435
DB 1442 AGTTTCTCTGCAAAAGCAGTGGACCATCAGGAAACCCAGTTCGCTATAGCAGCAGCG 1501
QY 1436 AAGTTAATCACTGAGTCCAAAGAGAGTACCACAGTGTGTCAGGCTGACTCTGACAGAT 1495
DB 1502 AAGTTAATCACTGAGTCCAAAGAGAGTACCACAGTGTGTCAGGCTGACTCTGACAGAT 1561
QY 1496 ATGCTCAGCCACTGGTAGGAGGATTTGTTGTTGATCACTTCAATCAAGATCTACCTTTAAAC 1555
DB 1562 ATGCTCAGCCACTGGTAGGAGGATTTGTTGTTGATCACTTCAATCAAGATCTACCTTTAAAC 1621
QY 1556 CAGAAGAGGAAAGAAAGAGCAGGCTATGCAGACCTAGATCTTACAACTCACCAGGCGAGG 1615
DB 1622 CAGAAGAGGAAAGAAAGAGCAGGCTATGCAGACCTAGATCTTACAACTCACCAGGCGAGG 1681
QY 1616 AAGTTTATCATGCTGCTGAACCACTCCCAATTAACGGGGCTGAGTATGCAACCCCAA 1675
DB 1682 AAGTTTATCATGCTGCTGAACCACTCCCAATTAACGGGGCTGAGTATGCAACCCCAA 1741
QY 1676 TCATCATGGACATGTCAGGGCACCCACACTTTCAGTTGGTTCAGCCCTCCACATCCACTT 1735
DB 1742 TCATCATGGACATGTCAGGGCACCCACACTTTCAGTTGGTTCAGCCCTCCACATCCACTT 1801
QY 1736 TCAAGGCTACGGGAAACCAACTCCCCACTAGTGGGAACCTTACAATACACTTCTCTCCA 1795
DB 1802 TCAAGGCTACGGGAAACCAACTCCCCACTAGTGGGAACCTTACAATACACTTCTCTCCA 1861
QY 1796 GGAAGTACAGCTGCTCTCAGCCCGGCGCCAGTATGATACCCCGAAAGCTGGGAAGCCAG 1855

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 00:35:34 ; Search time 5835.06 Seconds
(without alignments)
11371.681 Million cell updates/sec

Title: US-10-060-830-1
Perfect score: 2280
Sequence: 1 ggcgcgcgcgcgcctgggc.....ttgaagatgatgctgcttt 2280

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2280	100.0	2939	9	AF387547	AF387547 Homo sapi
2	2277.4	99.9	6093	9	AB073146	AB073146 Homo sapi
3	2152.2	94.4	3858	9	BC029658	BC029658 Homo sapi
4	1696.2	74.4	2339	10	AF387549	AF387549 Rattus no
5	1693	74.3	2842	10	AF387548	AF387548 Mus muscu
6	1596.2	70.0	2020	9	AK056350	AK056350 Homo sapi
7	1496.2	65.6	1907	6	AX405795	AX405795 Sequence
8	1158.4	50.8	1388	9	HUMCUB1	D29810 Human mRNA
9	488.8	21.4	192247	2	AC013497	AC013497 Homo sapi
10	488.8	21.4	193623	9	AC106728	AC106728 Homo sapi
11	477.8	21.0	146515	2	AC025661	AC025661 Homo sapi
12	370.4	16.2	586	9	BC007117	BC007117 Homo sapi
13	364	16.0	203102	2	AC126975	AC126975 Rattus no
14	232.4	10.2	164216	2	AC091213	AC091213 Homo sapi
15	224.2	9.8	1620	6	AX118820	AX118820 Sequence
16	224.2	9.8	1761	6	AX118818	AX118818 Sequence
17	224.2	9.8	1768	6	AX118822	AX118822 Sequence
18	222.6	9.8	2547	9	AK095973	AK095973 Homo sapi
19	218	9.6	192247	2	AC013497	AC013497 Homo sapi
20	204	8.9	1464	6	AX118816	AX118816 Sequence
21	198.8	8.7	59871	2	AC101046	AC101046 Mus muscu
22	174.8	7.7	217	6	AX182030	AX182030 Sequence
23	157.4	6.9	61344	2	AC073443	AC073443 Homo sapi
24	146.2	6.4	675	9	HS332516	AJ332516 Homo sapi
25	146.2	6.4	35018	9	AC026303	AC026303 Homo sapi
26	146.2	6.4	164216	2	AC091213	AC091213 Homo sapi
27	135.4	5.9	146515	2	AC025661	AC025661 Homo sapi
28	135.2	5.9	59871	2	AC101046	AC101046 Mus muscu
29	131.2	5.8	61344	2	AC073443	AC073443 Homo sapi
30	130.4	5.7	203102	2	AC126975	AC126975 Rattus no
31	119	5.2	2846	10	BC026771	BC026771 Mus muscu
32	94.4	4.1	6895	4	BOVFACV2A	M81441 Bos taurus
33	94.4	4.1	6910	4	BOVFACV	M81440 Bos taurus
34	92.8	4.1	1895	4	S80643	S80643 Bos taurus
35	92.8	4.1	1965	4	BTPAS67PT	X91895 B.taurus mr
36	91.6	4.0	6585	10	MM052925	U52925 Mus musculu
37	91.6	4.0	7062	4	AF191308	AF191308 Sus scrof
38	89.8	3.9	1719	9	HSU70312	U70312 Homo sapien
39	89.8	3.9	2728	9	BC030828	BC030828 Homo sapi
40	89.8	3.9	4681	6	AX473349	AX473349 Sequence
41	88.8	3.9	1227	4	BTBP47PRO	Y11719 Bos taurus
42	88.2	3.9	6893	9	HUMFVA	M14335 Human coagu
43	88	3.9	2303	10	AF031524	AF031524 Mus muscu
44	86.6	3.8	6909	6	A46255	A46255 Sequence 1
45	86.6	3.8	6909	6	A63218	A63218 Sequence 1

ALIGNMENTS

RESULT 1
AF387547

LOCUS

DEFINITION

Homo sapiens endothelial and smooth muscle cell-derived

neuropilin-like protein (ESDN) mRNA, complete cds.

ACCESSION AF387547

VERSION AF387547.1

KEYWORDS GI:16902434

SOURCE Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 2939)

Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,

Matsumori, A., Sasayama, S., Honjo, T. and Tashiro, K.

2939 bp mRNA linear PRI 11-NOV-2001


```

BC029658      3858 bp      mRNA      linear      PRI 20-MAY-2002
LOCUS
DEFINITION    Homo sapiens, similar to endothelial and smooth muscle cell-derived
               neuropeptide-like protein, clone MGC:30023 IMAGE:4431275, mRNA,
               complete cds.
ACCESSION     BC029658
VERSION       BC029658.1 GI:20988614
KEYWORDS      MGC.
SOURCE        Homo sapiens.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 3858)
AUTHORS       Strausberg, R.
TITLE         Direct Submission
JOURNAL       Submitted (06-MAY-2002) National Institutes of Health, Mammalian
               Gene Collection (MGC), Cancer Genomics Office, National Cancer
               Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
               USA
REMARK        NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT       Contact: MGC help desk
               Email: cgabs-z@mail.nih.gov
               Tissue procurement: DCTD/DTF
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
               DNA Sequencing by: Baylor College of Medicine Human Genome
               Sequencing Center
               Center code: BCM-HGSC
               Web site: http://www.hgsc.bcm.tmc.edu/cdna/
               Contact: amg@bcm.tmc.edu
               Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
               Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
               Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 42 Row: c Column: 3.
Location/Qualifiers
1..3858
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:30023 IMAGE:4431275"
/tissue_type="Prostate, adenocarcinoma."
/clone_lib="NIH_MGC_91"
/lab_host="DH10B"
/notes="vector: pCMV-SPORT6"
21..2252
/codon_start=1
/product="Similar to endothelial and smooth muscle
cell-derived neuropeptide-like protein"
/protein_id="AAH29658.1"
/db_xref="GI:20988615"
/translation="MPLFLLLLVLLLLLEDAGAQDGGCGHVLPGSGTSLTSINP
QTPNVTVCWEIRVMKGERVIRKFGDIEDSDCHFNLYRIYNGIGVSRTEIGRYC
GLGLQNMHSKESNGEITLLFMGSHVSGRFLASVLDKDLITCLDTASNFLRPE
FSKVPAGCLLPAEISGTPHGYRDSPLCMAGVHGVVNTLGGQISVVISKGIPI
YESLANNVTVVGHLSLFTFKTSGCCYGLMGESGVADPQITASSVLEWTHDYGQ
ENSWKPKAKLKPFPWAFATDEIQWLDLNKKEKITGITITGSTVWEHNYVSA
YRLISDDGQKWFIREPGEVDKIFGKNQDHOVNNFLPFIARFIRVNPTQWQ
KIAMELLECCQFIPKGRPPKLPQPPRNSNDLKNTPPKIAKGRAPKFTQLOPR
SSNEFPAQTEQTVASDPIRNTVTPNVTKDVALVLPVLMVLTLLILVCAWHW
RNRKKTQETVDLPYWDNRAGFYLMVSLACHRNEMWKMGPAPKAVDHEEPVRY
SSEVNLHSPREVTVLQADSAYEQPLVGVGLTHQSTFEKEEKEAGYADLDLYN
SPQEVTHAEPLTGPETPIIMDSGHEPTTSVGQSTSTFKATGNQPPPLVGT
YNTLLSRDSCSAQAQYDTPKAGKPLGAPDELVYQVPOSTQEVSGAGRDGCEVFK
EIL"
BASE COUNT    1137 a 807 c 785 g 1129 t
ORIGIN
Query Match      94.4%; Score 2152.2; DB 9; Length 3858;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 2212; Conservative 0; Mismatches 13; Indels 42; Gaps 1;

```


Qy	1082	GAATCTGTACAGTGTGATGGCAGAAATGGACATGTGTACAGAGAGCCCTGGTGTGGAGC	1141
Db	1449	GAGTCTGTGTACAGTGTGACATGGCAGAGATGGACATGTGTACAGAGAACCTGGTGTGGAGC	1508
Qy	1142	AAGATAAGATATTTCAAGGAACAAAGATATTACACAGGATGTGCGGTAACTAACTTTTTCG	1201
Db	1509	AGGACAAGATATTTCAAGGAACAAGATATTACAAGATGTTCTGTAATACTTTTTCG	1568
Qy	1202	CACCAATATTGCACGTTTTATTAGAGTGAATCTTACCCTAGTGCAGAGAAATTTGCCA	1261
Db	1569	CACCAATATTGCACGTTTCATTAGAGTGAACCTGTCCAGTGCACACAGAAATTTGCCA	1628
Qy	1262	TGAAATAGGAGTGTCTCGGATGTCAGTTTANTTCTTAAAGTGTCTCTCCAAAACTTTACTC	1321
Db	1629	TGAAAGTGGAACTGTCTCGGATGTCAAGTTTACTCTCAAGGTGCGCTTCCAAAGCTTTA---	1685
Qy	1322	AACCTCCACTCTCTGGAACAGCAATGACCTCAAAACACTACAGCCCTTCCAAAATAG	1381
Db	1686	---CTCACCTCTCTGGAACGCAATACCTTCAGATATCTACAGCTGTCTCCCAAACTAG	1742
Qy	1382	CCAAAGTGTGTGCCCCAAAATTTAGCGAACCACTACAACCTTCGCAGTAGCAATGAATTTTC	1441
Db	1743	GTAAAGTGTGTGCCCCATAATTTACTCAAGTGTCCAACCTCGAAGTAGGAATGAACCTTC	1802
Qy	1442	CTGCACAGACAGAAACAACTCCGCACTCTGTATATCAGAAATACTACCGTACACTCCAA	1501
Db	1803	CTGTGCAGCCGGGAGACAACTACCACTCTGTATATAAAAGAACACGACTGTAACTCCCA	1862
Qy	1502	ATGTAAACAAAGATGTAGCGCTGGCTGCAAGTTCTTGTCCCTGTCTGGTCAATGTCTCTCA	1561
Db	1863	GTGTAAACAAAGATGTGCACACTGGCTGCGCTTCTGGTCCCTGTGTGTCTATGGCCCTCA	1922
Qy	1562	CTACTCTCATTTCTCATATTAGTGTGTGTGTGGCACTGGAGAAACAGAAAGAAAAAACTG	1621
Db	1923	CCACACTCATCTTCATTCTAGTGTGTGTGGCACTGGAGAAACAGAAAGAAAAAACTG	1982
Qy	1622	AAGCCACTATGACTTACTTACTGGGACCGGGCAGTTTGTGGAAAGGAATGAAGCAGT	1681
Db	1983	AAGCGCTCTATGATTTACCCCACTGGGATCGGGCAGTTTGTGGAAAGGAATGAAGCAGC	2042
Qy	1682	TTCTTCTGCAAAAGCAGTGGACCATGTAGGAACCCCACTGTGGCTATAGCAGCAGCAAG	1741
Db	2043	TTCTCTCTGCCAAGTGGTGGACCAAGAGAGAGCCAGTGGCTTACAGCACTASTCTGAG	2102
Qy	1742	TTAATCACTGTAGTCCAAGAGAAGTCCACACAGTGTCTGCAGCTGACTCTCCAGAGTATG	1801
Db	2103	TCAGTCACTGTAGTGCAGGGAAGTCCACACAGTGTCTGCAGGCCGACTCTCCAGAAATG	2162
Qy	1802	CTCAGCACTGTGTAGGAGGAATTTGTGTGACTTCTATCAAGATCTACCTTTAAACAG	1861
Db	2163	CACAGCCCTCTGTGGGAGGAATTTGTGGCACACTCCATCAGAGATCCACCTTTAAACCTG	2222
Qy	1862	AAGAGGAAGAAAGCAGGCTATGCAGACTAGATCTTTACAACCTCACAGGCAGGAAG	1921
Db	2223	AGGAAGGAAGGAAGCAGGCTATGCAGACTCGATCTTTACAACCTCTCAATGCAGGAG	2282
Qy	1922	TTTATCATGCTCTGTCTGAACCACTCCCAATTTACGGGGCTTGAGTATGCAACCCCAATCA	1981
Db	2283	TGTACCAGCCTATGCTGTACCACTGCCGTATCAGGGGCTTGTGTACGCAACCCCGATCG	2342
Qy	1982	TCATGGACATGTACAGGACCCCAACACTTTCAGTTGGTCAGCCCTTCCACATCCACTTTCA	2041
Db	2343	TCATGGACATGTACAGGACCCCAACAGCCTCAGTTGGTGTGGAACTTACAACTCTTCTCTCCAGGA	2402
Qy	2042	AGGCTACGGGGAACCAACCTTCCCCACTAGTGGGAATTTACAATACACTTCTCTCCAGGA	2101
Db	2403	AAACTGTACGGGACCCCAACGCTTACGCTTTAGTGGGAATTTACAACACTTCTCTCTCCAGGA	2462
Qy	2102	CTGACAGCTGTCTTCAGCCCAAGCCCAAGTATGATACCCCGAAAGTGGGAAGCCAGGTC	2161
Db	2463	CTGACAGCTGTCTCTTCAGCCCAAGGCTCAGTATGACACCCCAAAAGGTGGGAAGTCAAGCT-	2521
Qy	2162	TACCTGCCCCACAGCAATTTGTGTACCAAGTGGCCACAGAGACACAAAGAAATATCAGGAG	2221

Db 2522 --GCTACCCAGAGAACTGGTATACCAAGGTGCCAGAGCAGCAGCTATCAGGAG 2579
 QY 2222 CAGGAAGGGATGGGAATGTGATGTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 2280
 Db 2580 CAGGAAGGGATGAGAAGTTTGATGCTTTTAAAGAAATCCTTTGACGGTGGTCTGCTTT 2638

RESULT 6
 AK056350
 LOCUS
 DEFINITION Homo sapiens cDNA FLJ31788 fis, clone NT2R12008598, weakly similar to MILK FAT GLOBULE-EGF FACTOR 8 PRECURSOR.
 ACCESSION AK056350
 VERSION AK056350.1 GI:16551729
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA, clone_lib:NT2R12 clone:NT2R12008598.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S., Ono,Y., Hotuta,T., Hirooka,S., Murakawa,K., Takiguchi,S., Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M., Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2020)
 AUTHORS Isogai,T., Otsuki,T. and Sugiyama,T.
 TITLE Direct Submission
 JOURNAL Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp; Tel:81-438-52-3975, Fax:81-438-52-3986)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- and 3'-end one pass sequencing: RAB HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES
 source Location/Qualifiers
 1. .2020
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2R12008598"
 /cell_line="NT2"
 /clone_lib="teratocarcinoma"
 /clone_lib="NT2R12"
 /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 5-weeks retinoic acid (RA) induction.-majorly NT2 neuron"
 BASE COUNT 494 a 519 c 530 g 477 t
 ORIGIN

Query Match 70.0%; Score 1596.2; DB 9; Length 2020;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1598; Conservative 0; Mismatches 3; Indels 0; Gaps 0

QY 1 GCCTGGCGCCCGCTGGCGGGCGCTCCCTCTCCGCGTCCCTCCCTGCTCCAC 60
 Db 420 GCCTGGCGCCCGCTGGCGGGCGCTCCCTCTCCGCGTCCCTCCCTGCTCCAC 479

QY 61 TCCTCCTCCTTCCTCCAGCTCTGTTCTCTCCTCTGCTCTTACTTGCTCTGCTGCTGCTC 120
 Db 480 TCCTCCTCCTTCCTCCAGCTCTGTTCTCTCCTCTGCTCTTACTTGCTCTGCTGCTGCTC 539

Db	1168	AGATAGATATTTCAAGGAAACAAAGAATATATACCAAGGATGGTGCCTATTAACCTTTT	12327
Qy	1200	GCACCAATATTATGCACGTTTTATTAGAGTGAATCCTACCCAATGGCAGCAAAATTGC	12599
Db	1228	GCACCAATATTATGCACG-TTTATTAGAGTGAATCCTACCCAATGGCAGCAAAATTGC	12866
Qy	1260	CATGAAATGGAGCTGCTCGGATGTCAGTTTATTCCTAAAGTGGTCTCTCCAAACCTTAC	13119
Db	1287	CATGAAATGGAGCTGCTCGGATGTCAGTTTATTCCTAAAGTGGTCTCTCCAAACCTTAC	13466
Qy	1320	TCAACCTCCACCTCCCTCGGAACAGCAATGACCTCAAAACACTACAGCGCCCTCCAAAAA-	13787
Db	1347	TCAACCTCCACCTCCCTCGGAACAGCAATGACCTC-AAAACACTACAGCGCCCTCCAAAAA	1405
Qy	1379	--TAGCCAAAGGTGTCGCCCAAAATTTACGCAACCACTACAACTCCGCTAGCAATGA	14366
Db	1406	TTTAGCCAAAGGTGTCGCCCAAAATTTACGCAACCACTACAACTCCGCTAGCAATGA	14655
Qy	1437	ATTTCCTGCACAGACAGAAACAACAACCTGCCAGTCTCTGATATCAGAAATACTACCGTAAC	14966
Db	1466	ATTTCCTGCACAGACAGACAAACAACACTGCCAGTCTCTGATATCAGAAATACTACCGTAAC	15255
Qy	1497	TCCAATGTAAACAAAGATGTAGCGGTGGCTGCAGTCTTTGTCCCTGTGCTGGTCATGGT	15566
Db	1526	TCCAATGTAAACAAAGATGTAGCGGTGGCTGCAGTCTTTGTCCCTGTGCTGGTCATGGT	15855
Qy	1557	CCTCAGTACTCTCATTTCTCATATTAGTGTGCTTGGCACTCGGAGAAACAGAAAGAAAA	16166
Db	1586	CCTCAGTACTCTCATTTCTCATATTAGTGTGCTTGGCACTCGGAGAAACAGAAAGAAAA	16455
Qy	1617	AACTGAAGGCACCTATGACTTACTCTTACTGGGACGGGCAGGT	1659
Db	1646	AACTGAAGGCACCTATGACTTACTCTTACTGGGACGGGCAGGT	1688

Qy	606	TGGAACAATTCTCTCATGGATATAGAGATTCTTCGCCATTGTGCATGCTGGTGTGCATGC	665
Db	631	TGGAACAATTCTCTCATGGATATAGAGATTCTTCGCCATTGTGCATGCTGGTGTGCATGC	690
Qy	666	AGGAGTAGTCTCAACACACGTTGGGCGGCCAATCAGTCTTGTAAATTAGTAAAGTATTCC	725
Db	691	AGGAGTAGTCTCAACACACGTTGGGCGGCCAATCAGTCTTGTAAATTAGTAAAGTATTCC	750
Qy	726	CTATTATGAAGTCTTTTGGCTAAACAAGTCACATCTCTGGTGGGACACATTATCTACAAG	765
Db	751	CTATTATGAAGTCTTTTGGCTAAACAAGTCACATCTCTGGTGGGACACATTATCTACAAG	810
Qy	786	TCTTTTTACATTTAAAGACAAGTGGATGTTATGGAAACCTGGGGATGGAGTCTGGT - GTGA	844
Db	811	TCTTTTTACATTTAAGACAAGTGGATGTTATGGAAACCTGGGGATGGAGTCTGGTGGTGA	870
Qy	845	TCGGGATGCTCAAAATAACAGCATCATCTGCTGGAGTGACCTGACCCACACAGGGCAAG	904
Db	871	TCGGGGATGCTCAAAATAACAGCATCA - CTGTCTGGAGTGACCTGACCCACACAGGGCAAG	929
Qy	905	AGACAGCTTGGAAACCC - AAAAAGCCAGGCTGAAAAAACCTGGACCGCCTTGGGCTGCT	963
Db	930	AGACAGCTTGGAAACCCAAAAAAGCCAGGCTGAAAAAACCTGGACCGCCTTGGG - TGCT	988
Qy	964	TTTGCCACTGATGAATACCACTGGTTACAAATAGATTTGAATAAGGAAAAGAAAATAACA	1023
Db	989	TTTGCCACTGATGAATACCACTGGTTACAAATAGATTTGAATAAGGAAAAGAAAATAACA	1048
Qy	1024	GGCATTAATACCACTGGATCCACCATGGTGGAGCACAAATTACTATGTCTGTGCCTACAGA	1083
Db	1049	GGCATTAATACCACTGGATCCACCATGGT - GAGCACAAATTACTATGTCTGTGCCTACAGA	1107
Qy	1084	ATCCTGTACAGTGATGATGGCCAGAAATGGACTGTGTACAGAGAGCCTGGT - GTGGAGCA	1142
Db	1108	ATCCTGTACAGTGATGATGGCCAGAAATGGACTGTGTACAGAGAGCCTGGTGTGGAGCA	1167
Qy	1143	AGATAAGNATTTTCAAGGAACAAA - GATTATCATCCAGGAT - GTGCGTAAATAACTTTTT	1199

	/note="assembly_fragment"	
misc_feature	30657..49294	
misc_feature	/note="assembly_fragment"	
misc_feature	49395..64509	
misc_feature	/note="assembly_fragment"	
misc_feature	64610..83042	
	/note="assembly_fragment"	
	clone_end:r7	
	vector_side:left	
misc_feature	83143..112373	
misc_feature	/note="assembly_fragment"	
misc_feature	112474..145663	
misc_feature	/note="assembly_fragment"	
misc_feature	145764..192247	
misc_feature	/note="assembly_fragment"	
BASE COUNT	58524 a 38157 c 36840 g 57721 t 1005 others	
ORIGIN		
Query Match	21.48; Score 488.8; DB 2; Length 192247;	
Best Local Similarity	99.6%; Pred. No. 1.6e-127;	
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
QY 1789	TCTGCAGATGTCACGCCACTGGTAGAGGAATTGTGGTACACTTCATCAAGAATCT 1848	
Db 164971	TCACAGATGTCACGCCACTGGTAGAGGAATTGTGGTACACTTCATCAAGAATCT 165030	
QY 1849	ACCTTTAAACCAGAAAGAGGAAAAGCAGCGCTATGCAGACCTAGATCCTTACAACCTCA 1908	
Db 165031	ACCTTTAACCAGAAAGAGGAAAAGCAGCGCTATGCAGACCTAGATCCTTACAACCTCA 165090	
QY 1909	CCAGGCGAGGAAGTTTTATCATGCTTGCTGAACCACTCCCAATTACGGGGCCTGAGTAT 1968	
Db 165091	CCAGGCGAGGAAGTTTTATCATGCTTGCTGAACCACTCCCAATTACGGGGCCTGAGTAT 165150	
QY 1969	GCAACCCCATACTCATGACATGTGCAGGCAACCCCAACTTCAGTTGTCAGCCCTCC 2028	
Db 165151	GCAACCCCATACTCATGACATGTGCAGGCAACCCCAACTTCAGTTGTCAGCCCTCC 165210	
QY 2029	ACATCCACATTTCAAGGCTACGGGGAACCAACCTCCCCCACTAGTGGGAACCTTACAATACA 2088	
Db 165211	ACATCCACATTTCAAGGCTACGGGGAACCAACCTCCCCCACTAGTGGGAACCTTACAATACA 165270	
QY 2089	CTTCTCTCAGAGCTGACAGCTGCTCTCAGCCAGGCCCAAGTATGATACCCCGAAGCT 2148	
Db 165271	CTTCTCTCAGAGCTGACAGCTGCTCTCAGCCAGGCCCAAGTATGATACCCCGAAGCT 165330	
QY 2149	GGGAACCCAGGTCTACCTGCCCCAGACGAATTTGTTGTAACCAAGTGCACAGACACACAA 2208	
Db 165331	GGGAACCCAGGTCTACCTGCCCCAGACGAATTTGTTGTAACCAAGTGCACAGACACACAA 165390	
QY 2209	GAAGTATCAGGACGAGGAAGGATGGGAATGTGATGTTTTAAAGAAATCCCTTTGAAGA 2268	
Db 165391	GAAGTATCAGGACGAGGAAGGATGGGAATGTGATGTTTTAAAGAAATCCCTTTGAAGA 165450	
QY 2269	TGATGCTGCTTT 2280	
Db 165451	TGATGCTGCTTT 165462	
RESULT 10		
AC106728		
LOCUS	Homo sapiens 3 BAC Rp11-319J24 (Roswell Park Cancer Institute Human	
DEFINITION	AC106728 193623 bp DNA linear PRI 25-MAY-2001	
ACCESSION	BAC Library) complete sequence.	
VERSION	AC106728.6 GI:21206081	
KEYWORDS	HTG.	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 193623)	
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R. Allen,C.,	

```

repeat_region      /rpt_family="MIR"
15395..15562
/rpt_family="MERSB"
16109..16388
/rpt_family="L1MC5"
16389..16642
/rpt_family="AluY"
16643..16758
/rpt_family="L1MC5"
16826..16898
/rpt_family="L1MC5"
/rpt_family="L1MC4"
18655)
complement(19191..19653)
/rpt_family="L1MD3"
19780)
complement(19662..19780)
/rpt_family="FLAM_A"
21049)
complement(19781..21049)
/rpt_family="L1MD3"
21455)
complement(21069..21455)
/rpt_family="WLR1B"
22318)
complement(21479..22318)
/rpt_family="L1MD3"
22681)
complement(22447..22681)
/rpt_family="L1MD3"
23853)
complement(23470..23853)
/rpt_family="WSTD"
23925)
complement(23479..23925)
/rpt_family="U3"
23936..25203
/rpt_family="L1PB3"
24844..24975
/standard_name="6234"
25219..25388
/rpt_family="TTTC)n"
25548)
complement(25428..25548)
/rpt_family="AluCb"
25950)
complement(25951..26456)
/rpt_family="L1PB3"
26457..27565
/rpt_family="MER75"

Query Match      21.4%; Score 488.8; DB 9; Length 193623;
Best Local Similarity 99.6%; Pred No. 1.6e-127;
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1789 TCTGCAGAGTATGCTCAGCCACTGTTAGAGGAATGTTGGTACACTTCATCAAGATCT 1848
|| |||||
DB 96184 TCCACAGAGTATGCTCAGCCACTGTTAGAGGAATGTTGGTACACTTCATCAAGATCT 96243

QY 1849 ACCTTTAACCCAGACAGGAAAGAACAGCAGCTATGCGACCTAGATCCTTACAACTCA 1908
|||||
DB 96244 ACCTTTAACCCAGACAGGAAAGAACAGCAGCTATGCGACCTAGATCCTTACAACTCA 96303

QY 1909 CCAGGCGAGAGTTTATCATGCTATGCTGAACCACTCCCAATTCAGGGGCTCAGTAT 1968
|||||
DB 96304 CCAGGCGAGAGTTTATCATGCTATGCTGAACCACTCCCAATTCAGGGGCTCAGTAT 96363

QY 1969 GCAACCCCAATCATCATGACATGTGAGGGCCACCCACAACTTCAGTTGGTTCAGCCCTCC 2028
|||||
DB 96364 GCAACCCCAATCATCATGACATGTGAGGGCCACCCACAACTTCAGTTGGTTCAGCCCTCC 96423

QY 2029 ACATCCACTTCAAGGCTACGGGACCAACCTCCCTCCCACTAGTGGGAACCTTACAACTCA 2088
|||||
DB 96424 ACATCCACTTCAAGGCTACGGGACCAACCTCCCTCCCACTAGTGGGAACCTTACAACTCA 96483

QY 2089 CTTCCTCCAGGACTCAGACGTCTCCTCAGCCCGAGGCCCGAGTATACCCCGGAAGCT 2148
|||||
DB 96484 CTTCCTCCAGGACTCAGACGTCTCCTCAGCCCGAGGCCCGAGTATACCCCGGAAGCT 96543

QY 2149 GGAAGCCAGGTCTACCTGCCAGACGAATGGTGTACAGGTGCCAGAGCACACAA 2208
|||||

```

Db 96544 GGAAGCCAGGTCTACCTGCCAGACGAATGGTGTACAGGTGCCAGAGCACACAA 96603

QY 2209 GAAGTATCAGGACGAGGAGGATGGATGATGTTTAAAGAAATCCCTTTGAAGA 2268
 |||||

Db 96604 GAAGTATCAGGACGAGGAGGATGGATGATGTTTAAAGAAATCCCTTTGAAGA 96663
 |||||

QY 2269 TGATGCTGCTTT 2280
 |||||

Db 96664 TGATGCTGCTTT 96675
 |||||

RESULT 11
 AC025661
 LOCUS
 DEFINITION Homo sapiens chromosome 3 clone RP11-322A20 map 3, LOW-PASS
 SEQUENCE SAMPLING.
 AC025661
 AC025661.2 GI:7543781
 VERSION
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 146515)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 3, clone RP11-322A20
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 146515)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., Mienga,V., Morrow,J.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., O'Connor,T., O'Donnell,P.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (12-NOV-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 146515)
 REFERENCE
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., Mienga,V., Morrow,J.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., O'Connor,T., O'Donnell,P.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

```
Db 57265 TCCACAGAGTATGCTCAGCCACTGGTAGGAGGAATGTTGGTACACTTCATCAAGATCT 57324
Qy 1849 ACCTTTAAACAGAGAGGAAAGAGAGGCTATGCAGACCTTAGATCCTTACAACCTCA 1908
Db 57325 ACCTTTAAACAGAGAGGAAAGAGAGGCTATGCAGACCTTAGATCCTTACAACCTCA 57384
Qy 1909 CCAGGCGAGGAGTTATCATGCTCTATGCTGAACCACTCCCAATTCAGGGGCTGAGTAT 1968
Db 57395 CCAGGCGAGGAGTTATCATGCTCTATGCTGAACCACTCCCAATTCAGGGGCTGAGTAT 57444
Qy 1969 GCAACCCCAATCATCATGAGCATGTCAGGCGACCCCAACCTTCAGTTGTCAGCCCTCC 2028
Db 57445 GCAACCCCAATCATCATGAGCATGTCAGGCGACCCCAACCTTCAGTTGTCAGCCCTCC 57504
Qy 2029 ACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCACTAGTGGGAACCTTACAATACA 2088
Db 57505 ACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCACTAGTGGGAACCTTACAATACA 57564
Qy 2089 CTTCTCTCCAGGACTGACAGCTGCTCCTCAGCCCGGAGGAGTATGATACCCCGAAGCT 2148
Db 57565 CTTCTCTCCAGGACTGACAGCTGCTCCTCAGCCCGGAGGAGTATGATACCCCGAAGCT 57624
Qy 2149 GGGAGCCAGGCTCTACCTGCCCCAGAGCAATGTTGTACCAAGTGCCACAGAGCACACAA 2208
Db 57625 GGGAGCCAGGCTCTACCTGCCCCAGAGCAATGTTGTACCAAGTGCCACAGAGCACACAA 57684
Qy 2209 GAAGTATCAGGAGGAGGAGGAGGAGGAGTGT-CATGTTTTTAAAGAAATCCTTTGAAG 2267
Db 57685 GAAGTATCAGGAGGAGGAGGAGGAGGAGTGTGATGTTTTTAAAGAAATCCTTTGAAG 57744
Qy 2268 ATGATGCTGCTTT 2280
Db 57745 ATGATGCTGCTTT 57757

RESULT 12
BC007117
LOCUS BC007117 586 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:4333276, mRNA, partial cds.
ACCESSION BC007117
VERSION BC007117.1 GI:13938000
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Strausberg, R.
Submitted (30-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC/BDTD/DTP
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcdpaxil@stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 21 Row: n Column: 22
Location/Qualifiers
1..586
/organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/clone="IMAGE:4333276"
/cisue_type="Skin, melanotic melanoma, high MDR."
/clone_lib="NIH_MGC_62"
/lab_host="DH10B"
/Vector: pDNR-LIB"
<L..106
/codon_start=2
/product="Unknown (protein for IMAGE:4333276)"
/protein_id="AAH07117.1"
/db_xref="GI:13938001"
/translation="LLVLLLEDEAGAOQVMEARILTAGCAEGHLS"
BASE COUNT 190 a 106 c 127 g 163 t
ORIGIN
```

```
Query Match 16.2%; Score 370.4; DB 9; Length 586;
Best Local Similarity 97.2%; Pred. No. 3.9e-94;
Matches 377; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 140 AAGGTGATGATGTGGACACACTGTACTAGGCGCTGAGAGTGGAAACCTTACATCCATAA 199
Db 113 AAGGTGATGATGTGGACACACTGTACTAGGCGCTGAGAGTGGAAACCTTACATCCATAA 172
Qy 200 ACTACCACAGACCTATCCCAACAGCAGCTGTTGTGTAATGGGAGATCCGTTGAAGATGG 259
Db 173 ACTACCACAGACCTATCCCAACAGCAGCTGTTGTGTAATGGGAGATCCGTTGAAGATGG 232
Qy 260 GAGAGAGATTCGCATCAAAATTTGGTGACCTTTGACATTTGAAGATTCCTGATCTTGTCACT 319
Db 233 GAGAGAGATTCGCATCAAAATTTGGTGACCTTTGACATTTGAAGATTCCTGATCTTGTCACT 292
Qy 320 TTAATTACTTGAGAAATTTATAATGGAATTTGGAGTGCAGACAGCAACTGAAATAGGCAAAATACT 379
Db 293 TTAATTACTTGAGAAATTTATAATGGAATTTGGAGTGCAGACAGCAACTGAAATAGGCAAAATACT 352
Qy 380 GTGCTCTGGGGTTGCAAAATGAACCAATTCATTAATGAATCAAAAGGCAATGAATACATATGC 439
Db 353 GTGCTCTGGGGTTGCAAAATGAACCAATTCATTAATGAATCAAAAGGCAATGAATACATATGC 412
Qy 440 TGTTCATGATGGGAATCCATGTTCTCGGACGCGGATTTTGGCTCCTACATCTCTTTATAG 499
Db 413 TGTTCATGATGGGAATCCATGTTCTCGGACGCGGATTTTGGCTCCTACATCTCTTTATAG 472
Qy 500 ATAACAAGATCTTAATTACTTGTGGA 527
Db 473 ATAACAAGATCTTAATTACTTGTGGA 500
```

```
RESULT 13
AC126975/c
LOCUS AC126975 203102 bp DNA linear HTG 11-JUL-2002
DEFINITION Rattus norvegicus clone CH230-234N3, *** SEQUENCING IN PROGRESS
***, 74 unordered pieces.
ACCESSION AC126975
VERSION AC126975.1 GI:21731383
KEYWORDS HTG; HTGS. PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 203102)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbiera, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Bouch, J., Bowie, S., Brivela, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Din, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
```

```

* 70371 70470: gap of unknown length
* 70471 72795: contig of 2326 bp in length
* 72797 72895: gap of unknown length
* 72897 75155: contig of 2256 bp in length
* 75153 75252: gap of unknown length
* 75253 77649: contig of 2397 bp in length
* 77650 77749: gap of unknown length
* 77750 80695: contig of 2946 bp in length
* 80696 80795: gap of unknown length
* 80796 83797: contig of 3002 bp in length
* 83798 83897: gap of unknown length
* 83898 86763: contig of 2866 bp in length
* 86764 86863: gap of unknown length
* 86864 89892: contig of 3029 bp in length
* 89893 89992: contig of 2113 bp in length
* 89993 92105: contig of 2113 bp in length
* 92106 92205: gap of unknown length
* 92206 94330: contig of 2125 bp in length
* 94331 94430: gap of unknown length
* 94431 96835: contig of 2405 bp in length
* 96836 96935: gap of unknown length
* 96936 98553: contig of 2618 bp in length
* 98554 99654: gap of unknown length
* 99654 102601: contig of 2948 bp in length
* 102602 102701: gap of unknown length
* 102702 103556: contig of 2655 bp in length
* 103557 105357: gap of unknown length
* 105357 105457: gap of unknown length
* 105457 109434: contig of 3978 bp in length
* 109435 109534: gap of unknown length
* 109535 112973: contig of 3439 bp in length
* 112974 113073: gap of unknown length

Query Match      16.0%; Score 364; DB 2: Length 203102;
Best Local Similarity 85.2%; Pred. No. 5.9e-92;
Matches 419; Conservative 0; Mismatches 70; Indels 3; Gaps 1;

QY 1789 TCTGCAGACTATGCTCAGCACTGGTAGGAGGAATTGTTGTACACTTCACTCAAGATCT 1848
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180181 TCCACAGATATGCACAGCCCTCTGCGGAGGAATTGTTGGCACACTCCATCAGAGATCT 180122

QY 1849 ACCTTTAAACACAGAGAGAAAGACGAGCTATGACAGACTAGATCTTACAACTCA 1908
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180121 ACCTTTAAACCTGAAGAGAGAAAGACGAGCTACGCGATCTAGACCCCTTACAACGCT 180062

QY 1909 CAGGCGCAGAGATTATATCATGCTATGCTGAACCACTCCCAATTACGGGGCTCAGTAT 1968
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180061 CCAGTACAGGAAGTGTATCATGCTACGCTAGCGGCTGCCGGTAAACGGGGCTCAGTAC 180002

QY 1969 GCAACCCCAATCATATGACATGTGCAGGCGACCCACCAACTTTCAGTTGGTCAGCCCTCC 2028
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180001 GCANCCCCAATCTCATGACATGTGCAGGCGACTCCACAGCTCAGTTGGTCTGCCCTCC 179942

QY 2029 ACATCCACTTTCAAGGCTACGGGGAACCACTCCCTCCCACTAGTGGGAATTTACAAATACA 2088
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179941 ACATCCACTTTCAAGACTGCAGGACCACTCCCTCCCGATTAGTGGGAATTTACAACT 179882

QY 2089 CTTCTCTCAGGACTGACAGCTGCTCTCAGCCCGCCAGCCAGTATGATACCCGAAAGCT 2148
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179881 CTTCTCTCAGGACTGACAGCTGCTCTCAGGCGCAGGCTCAGTACGACACCCCAAGGT 179822

QY 2149 GGAAGCCAGGTCTACTCTCCCGACAGGAATTGGTGTACAGGTGCCAGAGCACACAA 2208
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179821 GGAAGCCAG--CAGTCTCCCGACAGGAATGGTGTACAGGTGCCAGAGCACCCAG 179765

QY 2209 GAAGATCAGGACGAGGAGGATGGGAATGTGATGTTTTTAAAGAAATTCCTTTGAAGA 2268
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179764 GAAGCATCAGGACGAGGAGGATGAGAAATTTGATGCTTTTAAAGAAACCTTTGACGG 179705

QY 2269 TGATGCTGCTTT 2280
    ||||| ||||| |||||
Db 179704 TGGTGTGCTTT 179693

```

AC091213 164216 bp DNA linear HTG 24-JUN-2002
 Homo sapiens chromosome 3 clone CTD-2011L5, WORKING DRAFT SEQUENCE,
 8 unordered pieces.
 AC091213 GI:21539673
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 164216)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lousseg,H.,
 Lozadó,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okunolu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 River,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scher,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
 Sodergren,E., Sonaake,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

TITLE
 JOURNAL
 REFERENCE
 2 (bases 1 to 164216)
 Worley,K.C.
 Direct Submission
 Submitted (04-APR-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 164216)
 Worley,K.C.
 Direct Submission
 Submitted (24-JUN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jun 23, 2002 this sequence version replaced gi:21539104.
 ----- Genome Center of Medicine
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Drafting Center Code: BCM

```
Db 797 AAATCAGAGCTTCTTCCTCATGGCAGTCGGTCAATGAGAGTGGAGACCAAGTTCACCTGGT 856
Qy 917 AACCCAAAAGCCAGGCTGAAAAAACCCTGGACCGCCTTGGGCTGGCTTTTGGCCACTGATG 976
Db 857 CTCCTGGCCAAAGCCCGACTTCAGGACCAAGCCCATCATGGCTTCGGGGCGACAGTAGCA 916
Qy 977 AATACC-----AGTGGTTACAATAGATTTGAATAAGGAAAAGAAATAACAG 1024
Db 917 ACAACCAAAACCAGGAGTGGCTGGAGATCGATTTGGGGGAGAAAAGAAAATAACAG 976
Qy 1025 GCATTATAACCACTGGATCCACCATGGTGGAGCACAAATTAATATGTCTGCTCCCTACAGAA 1084
Db 977 GAATTAGCACACAGGATCTACACAGTCGAACTTCACTTTTATGTTAAGAGTTTTGTGA 1036
Qy 1085 TCCTGTACAGTGAATGGGCGAGAAATGGACTGTGTACAGAGAGCCCTGGTGTGGAGCAAG 1144
Db 1037 TGAACCTTCAAAAACAATAATTCTAAGTGGAAAGACCTATAAAGGAATTGTGAATAATGAAG 1096
Qy 1145 ATAAGATATTTCAAGGAACAAGATTTATCACCAGGATGTGCGTAAATAACTTTTGGCCAC 1204
Db 1097 AAAAGGTGTTTCAGGGTAACCTTAACCTTTCGGGACCCAGTGCNAAACAAATTCATCCCTC 1156
Qy 1205 CAATTATTGCCAGCTTTTATTAGAGTGAATCCTACCAATGGCAGCAGAAAATTCGCATGA 1264
Db 1157 CCATCGTGGCCAGATATGTGGGGTTGTCCCCCAGACATGGCACCAAGGATAGCCTTGA 1216
Qy 1265 AAATGGAGCTGCTCGGATGTCAGTTTA 1291
Db 1217 AGGTGGAGCTCATTGGTTGCCAGATTA 1243
```

Search completed: January 21, 2003, 07:01:32
Job time : 6530.06 secs

Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HClB
 Center clone name: CTD-2011L5
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 161349 bases at least Q40
 Consensus quality: 162327 bases at least Q30
 Consensus quality: 162604 bases at least Q20
 Estimated insert size: 157933; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 3452: contig of 3452 bp in length
 * 3453 3452: gap of unknown length
 * 3553 10549: contig of 6997 bp in length
 * 10550 10649: gap of unknown length
 * 10650 26000: contig of 15351 bp in length
 * 26001 26100: gap of unknown length
 * 26101 44000: contig of 17900 bp in length
 * 44001 44100: gap of unknown length
 * 44101 58020: contig of 13920 bp in length
 * 58021 58120: gap of unknown length
 * 58121 79695: contig of 21575 bp in length
 * 79696 119936: gap of unknown length
 * 119937 120036: contig of 40141 bp in length
 * 120037 164216: contig of 44180 bp in length.
 *
 * Location/Qualifiers
 source
 1..164216
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="CTD-2011L5"
 BASE COUNT 52340 a 31466 c 30520 g 49178 t 712 others
 ORIGIN
 Query Match 10.2%; Score 232.4; DB 2; Length 164216;
 Best Local Similarity 99.6%; Pred. No. 2.5e-54;
 Matches 233; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 140 AAGTGATGATGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTTACATCCATAA 199
 Db 31855 AAGTGATGATGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTTACATCCATAA 31914
 QY 200 ACTACCACAGACCTATCCCAACAGCACTGTTGTGAATGGGAGATCCGTTAAAGATGG 259
 Db 31915 ACTACCACAGACCTATCCCAACAGCACTGTTGTGAATGGGAGATCCGTTAAAGATGG 31974
 QY 260 GAGAGAGTTCGATCAAAATTTGGTACCTTGACATTTGAGATTCGTATCTGTGCAT 319
 Db 31975 GAGAGAGTTCGATCAAAATTTGGTACCTTGACATTTGAGATTCGTATCTGTGCAT 32034
 QY 320 TTAATTACTTGCAGAAATTTAATGGAATTTGAGTCAACAGCACTGAATATAGGCA 373
 Db 32035 TTAATTACTTGCAGAAATTTAATGGAATTTGAGTCAACAGCACTGAATATAGGTA 32088
 RESULT 15
 AX118820 1620 bp DNA linear PAT 11-MAY-2001
 LOCUS
 DEFINITION Sequence 5 from Patent WO0129219.

AX118820
 AX118820.1 GI:14035776
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 human.
 Mammalia; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1620)
 Donoho, G., Turner, C.A., Nehls, M., Friedrich, G., Zambrowicz, B. and
 Sands, A.T.
 Human cub-domain-containing protein and gene encoding the same
 Patent: WO 0129219-A 5 26-APR-2001;
 Lexicon Genetics Incorporated (US)
 FEATURES
 Location/Qualifiers
 source
 1..1620
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 456 a 353 c 422 g 389 t
 ORIGIN
 Query Match 9.8%; Score 224.2; DB 6; Length 1620;
 Best Local Similarity 52.0%; Pred. No. 2.6e-52;
 Matches 607; Conservative 0; Mismatches 518; Indels 42; Gaps 3;
 QY 137 AGCAAGGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTTACATCCA 196
 Db 107 AGCTGGGTGATGGCTGTGGACACCTAGTACTTATCAGATAGTGGCACAATGACATCTA 166
 QY 197 TAACTACCCACAGACCTATCCCAACAGCACTGTTGTGAATGGGAGATCCGTTAAAGA 256
 Db 167 AGAATTATCCCGGGACCTACCCCAATCACACTGTTGGGAAAGACAATTACAGTACCAA 226
 QY 257 TGGGAGAGAGATTCCGATCAAAATTTGGTGCATTTGACATTTGACATTTGATTTCTTCTC 316
 Db 227 AGGGGAAAGAACTGATCTGAGGTTGGGAGATTTGGATATCGAA----- 270
 QY 317 ACTTTAATTACTTGTGAGAATTTAATGGAATTTGAGTGCAGCAGAACTGAAATGAGCAAA 376
 Db 271 --TCCAGACCTGCTCTCTGACTATCTCTTCCACAGCTCTTCAGATCAATATGTC 328
 QY 377 ACTGTGGTCTGGGTTGCAAAATGAACCAATCAATTAATGAATCAAAAGGCAATGAATCAAT 436
 Db 329 CATACTGTGGAAGTATGACTGTTCCTCAAGAACTCTTGTGAACACACAGTGAAGTAAACG 388
 QY 437 TCCTGTTTCATGAGTGAATCCCATGTTCTCGAGCGGGATTTTGGCCTCATACTCTGTTA 496
 Db 389 TCCGCTTTGAGAGTGGATCCCACTTTCTGCGCGGGGTTTTTGTCTGACCTATGCGAGCA 448
 QY 497 TAGATAAACAAGATCTAATTACTTTGTCACACTGCATCCCAATTTTGGAACTGAGT 556
 Db 449 GCGACCATCCAGATTTAATAACATGTTTGGACGAGCTAGCCATTTATTTGAGACAGAA 508
 QY 557 TCAGTAAGTACTGCCAGCTGGTTGTCCTTCCCTTTGCTGAGATATCTGGAACAATTC 616
 Db 509 ACAGCAAAATCTGCCAGCTGGTTGTGAGAGACGTAGCAGGAGACATTTCTGGGAATATGG 568
 QY 617 CTCATGGATATAGAGATTCCTGCCATTTGTCATGGCTGGTGTGATGATGAGAGTAGTGT 576
 Db 569 TAGATGGATATAGAGATACCTCTTTATTTGTGCAAGCTGCCATCCATGAGGAATATGG 628
 QY 677 CAACACAGTTGGCGGCCCAAAATCAGTGTGTAATAGTAAGGATATTCCTATTATGAA 736
 Db 629 CTGATGAACATGAGTGGCCCAAGATCAGTGTGCTTCAGCGCAAGGGGATCAGTCGATATGAG 688
 QY 737 GTTCTTTGCTTAACAACAGTCACATCTGTGGGACACTTATCTACAAGCTCTTTTACAT 796
 Db 689 GGATTTCTGCCCAATGGTGTCTTTTCAGGGATGGTTCCCTGTCAGACACAGCGATTTCTGT 748
 QY 797 TTAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATCGCGGATCTCTC 856
 Db 749 TTACCTCCCAATGGTTTCAGCAGATCCTTGTAGTTTGAACCTGACGGG-----C 796
 QY 857 AATAACAGCATCATCTGTCTGGAGTGTGACTGTGACTGACCAACACAGCGGCAAGAGACAGTTGGA 916

Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loughheed, H., Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Ogun, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: G320
 Center clone name: CH230-234N3
 ----- Summary Statistics -----
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 145293 bases at least Q40
 Consensus quality: 154399 bases at least Q30
 Consensus quality: 161802 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 74 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1310: contig of 1310 bp in length
 1311: gap of unknown length
 1411: contig of 1087 bp in length
 2498: gap of unknown length
 2598: contig of 1121 bp in length
 3719: gap of unknown length
 3818: contig of 1156 bp in length
 4974: gap of unknown length
 5074: contig of 1044 bp in length
 4975: gap of unknown length
 5075: contig of 1044 bp in length
 6119: gap of unknown length

6219: contig of 1058 bp in length
 7277: gap of unknown length
 7377: contig of 1018 bp in length
 8395: gap of unknown length
 8495: contig of 1405 bp in length
 9900: gap of unknown length
 10000: contig of 1011 bp in length
 11010: gap of unknown length
 11111: contig of 1319 bp in length
 12429: gap of unknown length
 12530: contig of 1704 bp in length
 14233: gap of unknown length
 14234: contig of 1266 bp in length
 14334: gap of unknown length
 15600: contig of 1027 bp in length
 15700: gap of unknown length
 16727: contig of 2092 bp in length
 16826: gap of unknown length
 16918: contig of 1819 bp in length
 18919: gap of unknown length
 19019: contig of 1819 bp in length
 20838: gap of unknown length
 20937: contig of 2099 bp in length
 23037: gap of unknown length
 23136: contig of 1754 bp in length
 24890: gap of unknown length
 24891: contig of 1693 bp in length
 24931: gap of unknown length
 26884: contig of 1304 bp in length
 26784: gap of unknown length
 28088: contig of 1660 bp in length
 28187: gap of unknown length
 28188: contig of 1218 bp in length
 29847: gap of unknown length
 29848: contig of 1062 bp in length
 31165: gap of unknown length
 31265: contig of 1127 bp in length
 32327: gap of unknown length
 32427: gap of unknown length
 33554: contig of 1040 bp in length
 33555: gap of unknown length
 34695: contig of 1713 bp in length
 34794: gap of unknown length
 36507: contig of 1715 bp in length
 36508: gap of unknown length
 38322: contig of 1974 bp in length
 38423: gap of unknown length
 40397: contig of 2199 bp in length
 40496: gap of unknown length
 42695: contig of 1058 bp in length
 42795: gap of unknown length
 43853: contig of 1055 bp in length
 43953: gap of unknown length
 45008: contig of 1918 bp in length
 45109: gap of unknown length
 47026: contig of 2069 bp in length
 47127: gap of unknown length
 49195: contig of 2453 bp in length
 49295: gap of unknown length
 51748: contig of 2354 bp in length
 51849: gap of unknown length
 54202: contig of 2304 bp in length
 54303: gap of unknown length
 56606: contig of 1594 bp in length
 56707: gap of unknown length
 58300: contig of 1528 bp in length
 58401: gap of unknown length
 59928: contig of 2199 bp in length
 60028: gap of unknown length
 62227: contig of 2118 bp in length
 62228: gap of unknown length
 62328: contig of 2097 bp in length
 64446: gap of unknown length
 64546: contig of 1803 bp in length
 66643: gap of unknown length
 66742: contig of 1725 bp in length
 6743: gap of unknown length
 68545: contig of 1725 bp in length
 68546: gap of unknown length
 70370: contig of 1725 bp in length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 12, 2000 this sequence version replaced g1:7230271.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7942
 Center clone name: 322_A_20

* NOTE: This record contains 164 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 837 936: contig of 836 bp in length
 937 936: gap of 100 bp
 937 1748: contig of 812 bp in length
 1749 1848: gap of 100 bp
 1849 2680: contig of 832 bp in length
 2681 2780: gap of 100 bp
 2781 3534: contig of 754 bp in length
 3535 3634: gap of 100 bp
 3635 4383: contig of 749 bp in length
 4384 4483: gap of 100 bp
 4484 5261: contig of 778 bp in length
 5262 5361: gap of 100 bp
 5362 6136: contig of 775 bp in length
 6137 6236: gap of 100 bp
 6237 7045: contig of 809 bp in length
 7046 7145: gap of 100 bp
 7146 7953: contig of 808 bp in length
 7954 8053: gap of 100 bp
 8054 8857: contig of 804 bp in length
 8858 8957: gap of 100 bp
 8958 9784: contig of 827 bp in length
 9785 9884: gap of 100 bp
 9885 10718: contig of 834 bp in length
 10719 10818: gap of 100 bp
 10819 11640: contig of 822 bp in length
 11641 11740: gap of 100 bp
 11741 12534: contig of 794 bp in length
 12535 12634: gap of 100 bp
 12635 13407: contig of 773 bp in length
 13408 13507: gap of 100 bp
 13508 14276: contig of 769 bp in length
 14277 14376: gap of 100 bp
 14377 15185: contig of 809 bp in length
 15186 15285: gap of 100 bp
 15286 16086: contig of 801 bp in length
 16087 16186: gap of 100 bp
 16187 16986: contig of 800 bp in length
 16987 17086: gap of 100 bp
 17087 17913: contig of 827 bp in length
 17914 18013: gap of 100 bp
 18014 18841: contig of 828 bp in length
 18842 18941: gap of 100 bp
 18942 19770: contig of 829 bp in length
 19771 19870: gap of 100 bp
 19871 20689: contig of 819 bp in length

* 20690 20789: gap of 100 bp
 20790 21575: contig of 786 bp in length
 21576 21675: gap of 100 bp
 21676 22423: contig of 748 bp in length
 22424 22523: gap of 100 bp
 22524 23322: contig of 799 bp in length
 23323 23422: gap of 100 bp
 23423 24239: contig of 817 bp in length
 24240 24339: gap of 100 bp
 24340 25199: contig of 860 bp in length
 25200 25299: gap of 100 bp
 25300 26131: contig of 832 bp in length
 26132 26231: gap of 100 bp
 26232 27103: contig of 872 bp in length
 27104 27203: gap of 100 bp
 27204 28060: contig of 857 bp in length
 28061 28160: gap of 100 bp
 28161 29013: contig of 853 bp in length
 29014 29113: gap of 100 bp
 29114 29944: contig of 831 bp in length
 29945 30044: gap of 100 bp
 30045 30746: contig of 702 bp in length
 30747 30846: gap of 100 bp
 30847 31652: contig of 806 bp in length
 31653 31752: gap of 100 bp
 31753 32559: contig of 807 bp in length
 32560 32659: gap of 100 bp
 32660 33470: contig of 811 bp in length
 33471 33570: gap of 100 bp
 33571 34386: contig of 816 bp in length
 34387 34486: gap of 100 bp
 34487 35290: contig of 804 bp in length
 35291 35390: gap of 100 bp
 35391 36199: contig of 809 bp in length
 36200 36299: gap of 100 bp
 36300 37156: contig of 857 bp in length
 37157 37256: gap of 100 bp
 37257 38097: contig of 841 bp in length
 38098 38197: gap of 100 bp
 38198 39015: contig of 818 bp in length
 39016 39115: gap of 100 bp
 39116 39895: contig of 780 bp in length
 39896 39995: gap of 100 bp
 39996 40786: contig of 791 bp in length
 40787 40886: gap of 100 bp
 40887 41686: contig of 800 bp in length
 41687 41786: gap of 100 bp
 41787 42549: contig of 763 bp in length
 42550 42649: gap of 100 bp
 42650 43459: contig of 810 bp in length
 43460 43559: gap of 100 bp
 43560 44293: contig of 734 bp in length
 44294 44393: gap of 100 bp
 44394 45225: contig of 832 bp in length
 45226 45325: gap of 100 bp
 45326 46138: contig of 813 bp in length
 46139 46238: gap of 100 bp
 46239 46987: contig of 749 bp in length
 46988 47087: gap of 100 bp
 47088 47853: contig of 766 bp in length
 47854 47953: gap of 100 bp
 47954 48758: contig of 805 bp in length
 48759 48858: gap of 100 bp
 48859 49664: contig of 806 bp in length
 49665 49764: gap of 100 bp
 49765 50569: contig of 805 bp in length
 50570 50669: gap of 100 bp

Query Match 21.0%; Score 477.8; DB 2; Length 146515;
 Best Local Similarity 99.4%; Pred No. 2,1e-124;
 Matches 490; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1789 TCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTGTACACTTCATCAAGATCT 1848

Qy	121	GAGGAGCGCTGGAGCCCAAGTGATGGATGCGACACACTGCTACTAGCCCTCAGAGT	180
Db	540	GAGGAGCGCTGGAGCCCAAGTGATGGATGCGACACACTGCTACTAGCCCTCAGAGT	599
Qy	181	GGAAACCCCTACATCCATAAAGCTCCACAGACTATCCCAACAGCACTGTTGTGAATGG	240
Db	600	GGAAACCCCTACATCCATAAAGCTCCACAGACTATCCCAACAGCACTGTTGTGAATGG	659
Qy	241	GAGATCCGCTGTAAGATGGGAGAGAGAGTTCGGATCAAAATTTGGTGACTTTGACATTTGAA	300
Db	660	GAGATCCGCTGTAAGATGGGAGAGAGAGTTCGGATCAAAATTTGGTGACTTTGACATTTGAA	719
Qy	301	GATTCGTGATTCCTGCTACTTTAAATTAAGTGTGAGAAATTTATTAATGGAATTTGGAGTCAGCAGA	360
Db	720	GATTCGTGATTCCTGCTACTTTAAATTAAGTGTGAGAAATTTATTAATGGAATTTGGAGTCAGCAGA	779
Qy	361	ACTGAATAAGGCAAAATFACTGTGGTCTGGGGTTGCAAAATGAACCATTTCAATTTGAATCAAAA	420
Db	780	ACTGAATAAGGCAAAATFACTGTGGTCTGGGGTTGCAAAATGAACCATTTCAATTTGAATCAAAA	839
Qy	421	GGCAATGAATAACACATTCGCTGTCATGAGTGGAAATCCATGTTCTGGAGCGCGGATTTTGG	480
Db	840	GGCAATGAATAACACATTCGCTGTCATGAGTGGAAATCCATGTTCTGGAGCGCGGATTTTGG	899
Qy	481	GCCTCATCTCTGTTATAGATAAACAGATCAATTAATCTGTTTGGACACTGCATCCAAT	540
Db	900	GCCTCATCTCTGTTATAGATAAACAGATCAATTAATCTGTTTGGACACTGCATCCAAT	959
Qy	541	TTTTTTGGAACTTGAGTAAAGTACTGCCACACTGGTGTCTGCTTCCTTTTGGCTGAG	600
Db	960	TTTTTTGGAACTTGAGTAAAGTACTGCCACACTGGTGTCTGCTTCCTTTTGGCTGAG	1019
Qy	601	ATATCTGGAAACAAATTCCTCATGATATAGATTCCTCGCCATGTCGATGGCTGGTG	660
Db	1020	ATATCTGGAAACAAATTCCTCATGATATAGATTCCTCGCCATGTCGATGGCTGGTG	1079
Qy	661	CATCGAGGAGTATGTCCTAAACAGTGTGGCGGCCCAATTCAGTGTGTAATTAGTAAGGT	720
Db	1080	CATCGAGGAGTATGTCCTAAACAGTGTGGCGGCCCAATTCAGTGTGTAATTAGTAAGGT	1139
Qy	721	ATTCCCTATTATGAAGTCTTTTGGCTTAACAAGCTCAGATCTGTGTGGGACACTTATCT	780
Db	1140	ATTCCCTATTATGAAGTCTTTTGGCTTAACAAGCTCAGATCTGTGTGGGACACTTATCT	1199
Qy	781	ACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACATGGGGATGGAGTCTGGT	840
Db	1200	ACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACATGGGGATGGAGTCTGGT	1259
Qy	841	GTGATCGCGGATCCTCAAAATAACAGCATCATCTGTGCTGGAGTGGACTGACCAACACAGGG	900
Db	1260	GTGATCGCGGATCCTCAAAATAACAGCATCATCTGTGCTGGAGTGGACTGACCAACACAGGG	1319
Qy	901	CAAGAGAACAGTTGGAAACCCAAAAGCCAGGCTGAAAAACCTGGACCGCCTTGGGCT	960
Db	1320	CAAGAGAACAGTTGGAAACCCAAAAGCCAGGCTGAAAAACCTGGACCGCCTTGGGCT	1379
Qy	961	GCCTTTGCCACTCATGAATACCAAGTGGTTTACAAATAGATTTGAATAGGAAAAAGAAAATA	1020
Db	1380	GCCTTTGCCACTCATGAATACCAAGTGGTTTACAAATAGATTTGAATAGGAAAAAGAAAATA	1439
Qy	1021	ACAGGCATTATAACCACTGGATCCACATGTTGGAGACACAATTAATGCTGTGCTGCCTAC	1080
Db	1440	ACAGGCATTATAACCACTGGATCCACATGTTGGAGACACAATTAATGCTGTGCTGCCTAC	1499
Qy	1081	AGAAATCCTGTACAGTGAATGGCGAGAAATGGACTGTGTACAGAGACCTGGTGTGGAG	1140
Db	1500	AGAAATCCTGTGTGAGTGAATGGCGAGAAATGGACTGTGTACAGAGACCTGGTGTGGAG	1559
Qy	1141	CAAGATAGATATTTCAGGAAACAAAGATTATCACCAGGATGTCGTAATAACTTTTGG	1200
Db	1560	CAAGATAGATATTTCAGGAAACAAAGATTATCACCAGGATGTCGTAATAACTTTTGG	1619
Qy	1201	CCACCAATATTATGCAACGCTTTTATTAGATGAATCTCTACCAATGGCAGCAAAAATTGCC	1260

[illegible]

|||||
Db 1081 TGGAGCAAGATAAGATATTTCAAGGAACAAGATATATCACCAGAGTGTGCGTAATAACT 1140
|||||
QY 1196 TTTTCCACCAATATTCAGCTTTTATAGAGTGAATCCTACCAATGGCAGCAGAAAA 1255
|||||
Db 1141 TTTTCCACCAATATTCAGCTTTTATAGAGTGAATCCTACCAATGGCAGCAGAAAA 1200
|||||
QY 1256 TTGCCATGAAATGAGCTGCTCGATGTCAGTTTATTCCTTAAAGTGTCTCTCCAAAC 1315
|||||
Db 1201 TTGCCATGAAATGAGCTGCTCGATGTCAGTTTATTCCTTAAAGTGTCTCTCCAAAC 1260
|||||
QY 1316 TTACTCACTCCACCTCTCGGAACAGCAATGACCTCAAAACACATACACCCCTCCAA 1375
|||||
Db 1261 TTACTCACTCCACCTCTCGGAACAGCAATGACCTCAAAACACATACACCCCTCCAA 1320
|||||
QY 1376 AAATGCAAGAGTGTGCTGCCCAAAATTTAGCAACCACTCAACCTCGCAGTAGCAATG 1435
|||||
Db 1321 AAATGCAAGAGTGTGCTGCCCAAAATTTAGCAACCACTCAACCTCGCAGTAGCAATG 1380
|||||
QY 1436 AATTTCTGCGACAGACAGAAACAACACTGCCAGTCTGTATCATGAAATATACCCGTAA 1495
|||||
Db 1381 AATTTCTGCGACAGACAGAAACAACACTGCCAGTCTGTATCATGAAATATACCCGTAA 1440
|||||
QY 1496 CTCCTCAATGTAACCAAGATGTAGCGCTGGCTGAGTTCTTGTCTCTGCTGCTCATGG 1555
|||||
Db 1441 CTCCTCAATGTAACCAAGATGTAGCGCTGGCTGAGTTCTTGTCTCTGCTGCTCATGG 1500
|||||
QY 1556 TCTCACTACTCTCATTTCTCATTTAGTGTGCTGGCTGGCACTGGAGAAACAGAAAAA 1615
|||||
Db 1501 TCTCACTACTCTCATTTCTCATTTAGTGTGCTGGCTGGCACTGGAGAAACAGAAAAA 1560
|||||
QY 1616 AAACCTGAAGGCACTATGACTTACCTTACTGGGACCCGGC----- 1655
|||||
Db 1561 AAACCTGAAGGCACTATGACTTACCTTACTGGGACCCGGCAGGTTTATTTAATGTTCT 1620
|||||
QY 1656 -----AGTTGTGGAAGGAATGAAGCAGTTTCTTCTCTGCA 1693
|||||
Db 1621 CTTTGGCTGCGCTCACAAATGAAGTTGTGGAAGGAATGAAGCAGTTTCTTCTCTGCA 1680
|||||
QY 1694 AGCAGTGCACATGAGGAACCCAGTTCGCTATAGCAGCAGCAAGTATATCACTCA 1753
|||||
Db 1681 AGCAGTGCACATGAGGAACCCAGTTCGCTATAGCAGCAGTGAAGTATATCACTCA 1740
|||||
QY 1754 GTCCAGAGAAAGTACACAGAGTCTCGAGGCTGACTCTCGAGAGTATGCTCAGCCACTGG 1813
|||||
Db 1741 GTCCAGAGAGTACACAGAGTCTCGAGGCTGACTCTCGAGAGTATGCTCAGCCACTGG 1800
|||||
QY 1814 TAGAGGAATTTGTTGTACACTTCATCAAGATCTACCTTTAAACAGAGGAAGAAAG 1873
|||||
Db 1801 TAGAGGAATTTGTTGTACACTTCATCAAGATCTACCTTTAAACAGAGGAAGAAAG 1860
|||||
QY 1874 AAGCAGCTATGAGACCTAGATCCTTACAACTCAGCAGGAGGAAGTTTATCATGCCT 1933
|||||
Db 1861 AAGCAGCTATGAGACCTAGATCCTTACAACTCAGCAGGAGGAAGTTTATCATGCCT 1920
|||||
QY 1934 ATGCTGAACCACTCCCAATTTACGGGCTGAGTATGATCAACCCCAATCATCATGACATGT 1993
|||||
Db 1921 ATGCTGAACCACTCCCAATTTACGGGCTGAGTATGATCAACCCCAATCATCATGACATGT 1980
|||||
QY 1994 CAGGCAACCCCACTCACTTCACTGCTGAGTATGATCAACCCCAATCATCATGACATGT 2053
|||||
Db 1981 CAGGCAACCCCACTCACTTCACTGCTGAGTATGATCAACCCCAATCATCATGACATGT 2040
|||||
QY 2054 ACCAATCTCCCACTAGTGGGAACTTACAACTCTCTCCAGGACTGACAGCTGCT 2113
|||||
Db 2041 ACCAATCTCCCACTAGTGGGAACTTACAACTCTCTCCAGGACTGACAGCTGCT 2100
|||||
QY 2114 CCTCAGCCCGGCGGAGTATGATACCCGAAAGCTGGAGCCAGGTCTACCTGCCCCAG 2173
|||||
Db 2101 CCTCAGCCCGGCGGAGTATGATACCCGAAAGCTGGAGCCAGGTCTACCTGCCCCAG 2160
|||||
QY 2174 ACGAATTTGTTACAGGTCGCCACAGAGCACAAAGATATCAGGAGCAGGAGGATG 2233
|||||

Db 2161 ACGAATTTGTTACCAAGTCCACAGAGCACACAAGAGTATCAGGAGCAGGAGGATG 2220
QY 2234 GCGAATGTGATGTTTTTAAAGAAATCCTTTCAAGATGATGCTGCTTT 2280
|||||
Db 2221 GCGAATGTGATGTTTTTAAAGAAATCCTTTCAAGATGATGCTGCTTT 2267
|||||
RESULT 4
AF387549 2339 bp mRNA linear ROD 11-NOV-2001
LOCUS Rattus norvegicus endothelial and smooth muscle cell-derived
DEFINITION neurophilin-like protein (Esdn) mRNA, complete cds.
ACCESSION AF387549
VERSION AF387549.1 GI:16902438
KEYWORDS
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 2339)
Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,
Matsumori, A., Sasayama, S., Honjo, T. and Tashiro, K.
ESDN, a novel neurophilin-like membrane protein cloned from vascular
cells with the longest secretory signal sequence among eukaryotes,
is up-regulated after vascular injury (2001)
J. Biol. Chem. 276 (36), 34105-34114 (2001)
21424027
11447234
2 (bases 1 to 2339)
Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,
Sasayama, S., Matsumori, A., Honjo, T. and Tashiro, K.
Direct Submission
Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto
University Graduate School of Medicine, Yoshida Konoe-cho,
Sakyo-ku, Kyoto 606-8501, Japan
LOCATION/Qualifiers
1. .2339
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1. .2339
/gene="Esdn"
12. .2321
/gene="Esdn"
/product="endothelial and smooth muscle cell-derived
neurophilin-like protein"
/protein_id="AAL30180.1"
/db_xref="GI:16902439"
/translation="MASRAPLRAARSPDPGGRAPAAATGRAPLPSAGWCPLPPGRNS
SSRPLRLILLLLPDAGAOKGDCGHTVLGPESGTLTSINYPTYNSTVCKWEIRV
KTGRIRIKFGDFIEDSDYCHLNLYKIFNGIVSRTEIGYKGLQMNQSIKSGS
EITVLFMSGIHAGSGFLASVTDKODLITCLDTVSNFLPEFSKYCPACGLLPFAE
ISGTHPGYRSDSPCLCHAGIHAGVYSDVGGQISVYISKSTPYESSLANNVTSMVG
LSTSLFTKTSCTGLGMSGVADPOITASSVLENDHMGQENSKPEKARLEKPG
PPAAATFDEHQWLDLNEKKITGIVTSTLIEHNYVSAYRVLVYDDGQKWTVI
REPGAADKIFQKNKYHKNVNFLLP I IARFIRVNPVQOQKIAMKVELLGGQFTL
KGRLLPPTPPPPRNSNLKNTVHPKLGRAKFTQALQPRSRNDLPLPQTATPFD
VKNYTPSTVKDVALAVLPVLMALTTLLILVCAWHNRNKKKAEQYDLPFW
RAGWKGVKQLLPKAKSEHETPYRSNSEVSHLSPREVTTLQADSAEYAPLVGGI
VGLHQRTFKPEKEKASYADLPDYPAPVOEVHAYAEPLVPVGPVATVIVDMSC
HSTASVGLPSTFTRTAGNPPALVGYNTLLSRDTSQSGQAQIDTPKGGKPAAPAE
ELVYVQPSQTEASGARDKDFAKETL"
BASE COUNT 612 a 606 c 594 g 527 t
ORIGIN

Query Match 74.4%; Score 1696.2; DB 10; Length 2339;
Best Local Similarity 85.0%; Pred. No. 0;
Matches 1938; Conservative 0; Mismatches 323; Indels 18; Gaps 3;
QY 2 CCGCCGCCCGCGCTGGCGCGCTCCCTCTCCGCTCCCTCCCTCCCTCCCTCAACT 61
|||||

TITLE ESDN, a novel neuropilin-like membrane protein cloned from vascular cells with the longest secretory signal sequence among eukaryotes, is up-regulated after vascular injury

JOURNAL J. Biol. Chem. 276 (36), 34105-34114 (2001)

21424027 MEDLINE
11447234 PUBMED
2 (bases 1 to 2939)
Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,
Sasayama, S., Matsumori, A., Honjo, T. and Tashiro, K.

TITLE
JOURNAL
DIRECT SUBMISSION
Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 605-8501, Japan
Tsunetaka Otsuka

FEATURES	SOURCE
----------	--------

gene

CDS

/codon_start=1
 /product=-endothelial and smooth muscle cell-derived
 neuropeilin-like protein"
 /protein_id="AAL30178.1"
 /db_xref="GI:16902435"
 /translation="MASRVVRRRCPCQPVRAAAAAPAAALPLSRSLPPCSNSSSS
 FSPMLFLLVLLVLLLEDAGAQDGGCHTVLGPSTGLTISNYQTVYSPWCWE
 IRVYKREVIIRKFGDDIEDSDCHENYLIRVYNGISVTRTEICKYCGLGQHMNHSIES
 KNEITLLPMSGYTHVSRGFLASVYDKODLTCLDTSNLEFSPFKSYACGCLLP
 FAEISITLPHGSDYRSCMGVHAGVSNLTGGQISVILSKPIYKVESLIANNVSV
 VCHLSRLTEFKTSGCYGLMGSEYIAQPIYASSVLEHTDHTGQENSHKPKAKLK
 KQGPFAAATDIEYQMLDILNREKKITGLITTGSRVYENHYVSAIRLLISDQGCQ
 TVYRFGVEDDKIFQGNKQHQDVNRNLFPIIARFIRVNPYQVQAKRMKMLLQCG
 FIPKRAPPLQTTPPNRNSDLAKNTAPPKIAGRAPKFTQPLQPSRNSNFKPQETGT
 TASPDLINTVTPNVKNDAAAVLPVLMVLTLLILLCVCAWHNRNFKTEGTGTD
 LPVWDGVMWGMKMKLPKAPKADEETPVYRSGVENHLSPREVTTVLQDASAEYAP
 LKYGIVRGLTQGRSTFKPEAGVEYADLDPSNGQEVHYHAYEPLITGTPEVATPII
 MDMSGHPTTSFGQSTFKATGNQPPPLVGTYNLLSRLDTSQSSAAQYDTPKAGKP
 GLPAPDELVYVQVPQSTFVAGRGDCDVFKEIL"
 825 a 675 c 657 g 782 t
 BASE COUNT
 ORIGIN

Query Match	100.0%;	Score 2280;	DB 9;	Length 2939;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2280;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 1	CGCGCGCGCGCGCTGGGCGCGGCTCCCTCTCCCGCTCCCTCCCTCCCTGTCGTC	AAAC 60		
Db 75	CGCGCGCGCGCGCTGGGCGCGGCTCCCTCTCCCGCTCCCTCCCTGTCGTC	AAAC 134		
Qy 61	TCCTCCCTCTCTCCATGCCCTCTGTTCTCTCTGCTTACTTGCTGCTGCTGCTGTC	120		
Db 135	TCCTCCCTCTCTCTCCATGCCCTGTTCTCTGCTTACTTGCTGCTGCTGCTGTC	194		
Qy 121	GAGACGCTGGAGCCCGACGAGGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGT	180		
Db 195	GAGACGCTGGAGCCCGACGAGGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGT	254		
Qy 181	GGAAACCTTTACATCCATAAATACCCACAGACCTATCCCAACAGCACATGTTGTGAAATGG	240		
Db 255	GGAAACCTTTACATCCATAAATACCCACAGACCTATCCCAACAGCACATGTTGTGAAATGG	314		
Qy 241	GAGATCCCGTGTAAAGATGGGAGAGAGATTCCGATCAAAATTTGGTGACTTTGACATTGAA	300		
Db 315	GAGATCCCGTGTAAAGATGGGAGAGAGATTCCGATCAAAATTTGGTGACTTTGACATTGAA	374		
Qy 301	GATTCTGATCTCTGTCACTTTAAATTTACTTTGAGAATTTTAATGGAAATTTGGAGTCAGCAGA	360		
Db 375	GATTCTGATCTCTGTCACTTTAAATTTACTTTGAGAATTTTAATGGAAATTTGGAGTCAGCAGA	434		
Qy 361	ACTGAAATAGGCAAAATACTGTGGTCTGGGGTTGCAAAATGAACCAATTCATTAATGAAATCAAAA	420		

Db	435	ACTGAAATAGGCAAAATACTGTGCTCGGGGTGCAAAATGAACATTCAATTAATGAATCAAAA	494
Qy	421	GGCAATGAAATACATATGCTGTTTCATGAGTGAATCCATGTTTCTCGAGCGCGAATTTTG	480
Db	495	GGCAATGAAATACATATGCTGTTTCATGAGTGAATCCATGTTTCTCGAGCGCGAATTTTG	554
Qy	481	GCCTCATACTCTGTTATAGATAAACAAGATCTAATTAATTAATTTGTTGGACACTGCAATCAAT	540
Db	555	GCCTCATACTCTGTTATAGATAAACAAGATCTAATTAATTAATTTGTTGGACACTGCAATCAAT	614
Qy	541	TTTTTGGAACTGAGTTTCAGTAAGTACTGCCAGCTGGTTGTCGTCTCTTTTGGCTGAG	600
Db	615	TTTTTGGAACTGAGTTTCAGTAAGTACTGCCAGCTGGTTGTCGTCTCTTTTGGCTGAG	674
Qy	601	ATATCTGGAACAATTCCTCATGATATAGAGATCTCTCGCCATGTGTCATGGCTGGTGTG	660
Db	675	ATATCTGGAACAATTCCTCATGATATAGAGATCTCTCGCCATGTGTCATGGCTGGTGTG	734
Qy	661	CATCAGGAGTAGTGTCAAAACACGTTGGGCGCCAAATCAGTGTGTAATTAAGTAAAGT	720
Db	735	CATCAGGAGTAGTGTCAAAACACGTTGGGCGCCAAATCAGTGTGTAATTAAGTAAAGT	794
Qy	721	ATTCCCTATTATGAAAGTTCTTTGGCTAAACAACGTCACATCTGTGTGGGACACTTATCT	780
Db	795	ATTCCCTATTATGAAAGTTCTTTGGCTAAACAACGTCACATCTGTGTGGGACACTTATCT	854
Qy	781	ACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGTT	840
Db	855	ACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGTT	914
Qy	841	GTGATCGCGGATCCTCAAAATAACAGCATCATCTGCTGGAGTGGACTGACACACAGGG	900
Db	915	GTGATCGCGGATCCTCAAAATAACAGCATCATCTGCTGGAGTGGACTGACACACAGGG	974
Qy	901	CAAGAGAACAGTTGGAAACCCAAAAGCCAGGCTGAAAAAACCTGGACCGCCTTGGGCT	960
Db	975	CAAGAGAACAGTTGGAAACCCAAAAGCCAGGCTGAAAAAACCTGGACCGCCTTGGGCT	1033
Qy	961	GCTTTTGGCACTGATGAATACACAGTGGTTACAATACATTTGAATAGGAAAAAGAAATA	1022
Db	1035	GCTTTTGGCACTGATGAATACACAGTGGTTACAATACATTTGAATAGGAAAAAGAAATA	1099
Qy	1021	ACAGGCATTAATACCACTGGATCCACCATGGTGGAGCACAAATTAATATGTCTGCCTAC	1088
Db	1095	ACAGGCATTAATACCACTGGATCCACCATGGTGGAGCACAAATTAATATGTCTGCCTAC	1155
Qy	1081	AGATTCCTGTACAGTGATGATGGGCAGAAATGAGTCTGTACAGAGAGCCCTGGTGGAG	1114
Db	1155	AGATTCCTGTACAGTGATGATGGGCAGAAATGAGTCTGTACAGAGAGCCCTGGTGGAG	1212
Qy	1141	CAAGATAAGATATTCAAGGAACCAAGATTTATCACCAGGATGTGCGTAATAACTTTTGG	1200
Db	1215	CAAGATAAGATATTCAAGGAACCAAGATTTATCACCAGGATGTGCGTAATAACTTTTGG	1272
Qy	1201	CCACCAATTAATGACAGTCTTTATAGAGTGAATCCTACCCAAATGGCAGAGAAATTTGCC	1266
Db	1275	CCACCAATTAATGACAGTCTTTATAGAGTGAATCCTACCCAAATGGCAGAGAAATTTGCC	1333
Qy	1261	ATGAAAATGGAGTGTCTCGGATGTCAGTTTATTCCTTAAGAGTGTCTCTCCAAAACCTTACT	1322
Db	1335	ATGAAAATGGAGTGTCTCGGATGTCAGTTTATTCCTTAAGAGTGTCTCTCCAAAACCTTACT	1399
Qy	1321	CAACCTCCACCTCCTCGGAACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAATA	1388
Db	1395	CAACCTCCACCTCCTCGGAACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAATA	1455
Qy	1381	GCCTAAGTGTGTCGCCAAAATTTACGCAACCACTACAACCTCGCAGTAGCAATGAATTT	1444
Db	1455	GCCTAAGTGTGTCGCCAAAATTTACGCAACCACTACAACCTCGCAGTAGCAATGAATTT	1511
Qy	1441	CCTGCACAGACAGAAACAACACTGCCAGTCCCTGATATCAGAAATACTACCGTAACCTCA	1508

Db 451 TTTTAAAGAAATCCTT 467
|||||

RESULT 9
AAK51451
ID AAK51451 standard; DNA; 467 BP.
XX
AC AAK51451;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 26008.
XX
KW Human: bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 26008; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;

Query Match 93.5%; Score 257; DB 22; Length 467;
Best Local Similarity 100.0%; Pred. No. 2.3e-117;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAACCTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCACTCCCC 60
|||||
Db 211 CAACCTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCACTCCCC 270
|||||
Qy 61 CACTAGTGGGAACCTTACAAATACACTTCTCCAGGACTGACAGCTGCTCCTCAGCCAGG 120
|||||
Db 271 CACTAGTGGGAACCTTACAAATACACTTCTCCAGGACTGACAGCTGCTCCTCAGCCAGG 330
|||||
Qy 121 CCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCGACAGCAATTTGGTGT 180
|||||
Db 331 CCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCGACAGCAATTTGGTGT 390
|||||
Qy 181 ACCAGGTGCCACAGACACACAAGAATGATCAGGAGCAGGAAGGATGGGGAATGTGATG 240
|||||
Db 391 ACCAGGTGCCACAGACACACAAGAATGATCAGGAGCAGGAAGGATGGGGAATGTGATG 450
|||||

Qy 241 TTTTAAAGAAATCCTT 257
|||||
Db 451 TTTTAAAGAAATCCTT 467
|||||

RESULT 10
AAI57541
ID AAI57541 standard; DNA; 467 BP.
XX
AC AAI57541;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #26227 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 26227; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;

Query Match 93.5%; Score 257; DB 22; Length 467;
Best Local Similarity 100.0%; Pred. No. 2.3e-117;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAACCTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCACTCCCC 60
|||||
Db 211 CAACCTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCACTCCCC 270
|||||
Qy 61 CACTAGTGGGAACCTTACAAATACACTTCTCCAGGACTGACAGCTGCTCCTCAGCCAGG 120
|||||
Db 271 CACTAGTGGGAACCTTACAAATACACTTCTCCAGGACTGACAGCTGCTCCTCAGCCAGG 330
|||||
Qy 121 CCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCGACAGCAATTTGGTGT 180
|||||
Db 331 CCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCGACAGCAATTTGGTGT 390
|||||
Qy 181 ACCAGGTGCCACAGACACACAAGAATGATCAGGAGCAGGAAGGATGGGGAATGTGATG 240
|||||
Db 391 ACCAGGTGCCACAGACACACAAGAATGATCAGGAGCAGGAAGGATGGGGAATGTGATG 450
|||||

QY 241 TTTTAAAGAAATCCTT 257
 |||||
 DB 451 TTTTAAAGAAATCCTT 467

RESULT 11

AAI73442
 ID AAI73442 standard; DNA; 51 BP.

XX AC AAI73442;

XX DT 09-NOV-2001 (first entry)

XX DE Human silent SNP containing nucleic acid SEQ:383.

XX KW Human; single nucleotide polymorphism; SNP; genome; gene therapy;
 KW protein therapy; vaccine; probe; diagnostic assay; detection;
 KW quantitation; restorative therapy; polymorphic; ds.

XX OS Homo sapiens.

XX PN WO200140521-A2.

XX PD 07-JUN-2001.

XX PF 30-NOV-2000; 2000WO-US32758.

XX PR 30-NOV-1999; 99US-0168138.

XX PR 29-NOV-2000; 2000US-0726173.

XX PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach M;

XX WPI; 2001-356160/37.

XX PT Polymorphic nucleic acid sequences, useful in genetic testing and
 PT therapy -

XX PS Claim 1; Page 171; 2653pp; English.

XX CC AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
 CC sequences (I), which contain single nucleotide polymorphisms (SNPs).
 CC AAI73114 to AAI73329 represent peptides related to human polymorphic
 CC polynucleotide sequences. The sequences can be used in gene and protein
 CC therapy, and in vaccine production. (I) and the polypeptides encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate expression of polymorphic polypeptides.
 CC For example, (I) may be used to treat disorders by rectifying mutations
 CC or deletions in a patient's genome that affect the activity of
 CC polypeptides by expressing inactive proteins or to supplement the
 CC patients own production of polypeptide. Additionally, (I) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids
 CC in samples, and therefore which patients may be in need of restorative
 CC therapy. The polypeptides encoded by (I) may be used as antigens in the
 CC production of antibodies specific for polymorphic polypeptides. The
 CC antibodies may also be used to down regulate expression and activity.
 CC The antibodies may also be used as diagnostic agents for detecting the
 CC presence of polymorphic polypeptides in samples.

XX SQ Sequence 51 BP; 13 A; 17 C; 14 G; 7 T; 0 other;

Query Match 18.5%; Score 51; DB 22; Length 51;
 Best Local Similarity 100.0%; Pred. No. 2.2e-15;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 CAGCCAGCCAGTATGATATACCCGAAAGCTGGGAGCCAGGTCTACCTG 162
 |||||
 DB 1 CAGCCAGCCAGTATGATATACCCGAAAGCTGGGAGCCAGGTCTACCTG 51

RESULT 12

AAI73443
 ID AAI73443 standard; DNA; 51 BP.

XX AC AAI73443;

XX DT 09-NOV-2001 (first entry)

XX DE Human silent SNP containing nucleic acid SEQ:384.

XX KW Human; single nucleotide polymorphism; SNP; genome; gene therapy;
 KW protein therapy; vaccine; probe; diagnostic assay; detection;
 KW quantitation; restorative therapy; polymorphic; ds.

XX OS Homo sapiens.

XX PN WO200140521-A2.

XX PD 07-JUN-2001.

XX PF 30-NOV-2000; 2000WO-US32758.

XX PR 30-NOV-1999; 99US-0168138.

XX PR 29-NOV-2000; 2000US-0726173.

XX PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach M;

XX WPI; 2001-356160/37.

XX PT Polymorphic nucleic acid sequences, useful in genetic testing and
 PT therapy -

XX PS Claim 1; Page 172; 2653pp; English.

XX CC AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
 CC sequences (I), which contain single nucleotide polymorphisms (SNPs).
 CC AAI73114 to AAI73329 represent peptides related to human polymorphic
 CC polynucleotide sequences. The sequences can be used in gene and protein
 CC therapy, and in vaccine production. (I) and the polypeptides encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate expression of polymorphic polypeptides.
 CC For example, (I) may be used to treat disorders by rectifying mutations
 CC or deletions in a patient's genome that affect the activity of
 CC polypeptides by expressing inactive proteins or to supplement the
 CC patients own production of polypeptide. Additionally, (I) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids
 CC in samples, and therefore which patients may be in need of restorative
 CC therapy. The polypeptides encoded by (I) may be used as antigens in the
 CC production of antibodies specific for polymorphic polypeptides. The
 CC antibodies may also be used to down regulate expression and activity.
 CC The antibodies may also be used as diagnostic agents for detecting the
 CC presence of polymorphic polypeptides in samples.

XX SQ Sequence 51 BP; 14 A; 17 C; 13 G; 7 T; 0 other;

Query Match 9.1%; Score 25; DB 22; Length 51;
 Best Local Similarity 100.0%; Pred. No. 0.016;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 CAGCCAGCCAGTATGATATACCC 136
 |||||
 DB 1 CAGCCAGCCAGTATGATATACCC 25

RESULT 13

ABL19472/c
 ID: ABL19472 standard; DNA; 3751 BP.

XX ABL19472;

XX DT 26-MAR-2002 (first entry)

```
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 9889.
XX XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW Pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PD 23-MAR-2001; 2001WO-US09231.
XX PF 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX XX
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX XX
XX PS Claim 1; SEQ ID NO 9889; 21pp + Sequence Listing; English.
XX XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX SQ Sequence 3751 BP; 1086 A; 831 C; 790 G; 1044 T; 0 other;

Query Match          6.9%; Score 19; DB 23; Length 3751;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 TGATGTTTTTAAAGAAATC 254
      |||||||
Db 3277 TGATGTTTTTAAAGAAATC 3259

RESULT 14
AAK89089
ID AAK89089 standard; DNA; 9417 BP.
XX AC AAK89089;
XX XX
XX DT 05-NOV-2001 (first entry)
XX XX
XX DE Human digestive system antigen genomic sequence SEQ ID NO: 2665.
XX XX
XX KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
XX KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX KW digestive system disorder; Meckel's diverticulum; ds.
XX OS Homo sapiens.
XX XX
XX PN WO200155314-A2.
XX XX
XX PD 02-AUG-2001.
XX XX
XX PF 17-JAN-2001; 2001WO-US01324.

XX 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 11-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225477.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 18-AUG-2000; 2000US-0225759.
XX PR 22-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0226868.
XX PR 23-AUG-2000; 2000US-0227182.
XX PR 30-AUG-2000; 2000US-0227009.
XX PR 01-SEP-2000; 2000US-0228924.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229343.
XX PR 01-SEP-2000; 2000US-0229344.
XX PR 01-SEP-2000; 2000US-0229345.
XX PR 05-SEP-2000; 2000US-0229509.
XX PR 05-SEP-2000; 2000US-0229513.
XX PR 06-SEP-2000; 2000US-0230437.
XX PR 06-SEP-2000; 2000US-0230438.
XX PR 08-SEP-2000; 2000US-0231242.
XX PR 08-SEP-2000; 2000US-0231243.
XX PR 08-SEP-2000; 2000US-0231244.
XX PR 08-SEP-2000; 2000US-0231413.
XX PR 08-SEP-2000; 2000US-0231414.
XX PR 08-SEP-2000; 2000US-0232080.
XX PR 12-SEP-2000; 2000US-0232081.
XX PR 14-SEP-2000; 2000US-0232397.
XX PR 14-SEP-2000; 2000US-0232398.
XX PR 14-SEP-2000; 2000US-0232399.
XX PR 14-SEP-2000; 2000US-0232400.
XX PR 14-SEP-2000; 2000US-0232401.
XX PR 14-SEP-2000; 2000US-0233063.
XX PR 14-SEP-2000; 2000US-0233064.
XX PR 14-SEP-2000; 2000US-0233065.
XX PR 21-SEP-2000; 2000US-0234223.
XX PR 21-SEP-2000; 2000US-0234274.
XX PR 25-SEP-2000; 2000US-0234997.
XX PR 25-SEP-2000; 2000US-0234998.
XX PR 26-SEP-2000; 2000US-0235484.
XX PR 27-SEP-2000; 2000US-0235834.
XX PR 27-SEP-2000; 2000US-0235836.
XX PR 29-SEP-2000; 2000US-0236327.
XX PR 29-SEP-2000; 2000US-0236367.
XX PR 29-SEP-2000; 2000US-0236368.
XX PR 29-SEP-2000; 2000US-0236369.
XX PR 29-SEP-2000; 2000US-0236370.
```

PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0255719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-502630/55.

PT Polynucleotides encoding digestive system antigens, useful for
diagnosing, treating, preventing and/or prognosing disorders of the

PT digestive system, particularly cancer and cancer metastases -
XX Disclosure; SEQ ID NO 2665; 986pp; English..

XX The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.

XX SQ Sequence 9417 BP; 3043 A; 1599 C; 1757 G; 3012 T; 6 other;

Query Match 6.98; Score 19; DB 22; Length 9417;

Best Local Similarity 100.0%; Pred. NO. 14;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TCAGTTGGTCAGCCCTCCA 24

|||||

Db 4701 TCAGTTGGTCAGCCCTCCA 4719

RESULT 15

AAI57634
ID AAI57634 standard; DNA; 9417 BP.

XX AC AAI57634;

XX 19-OCT-2001 (first entry)

DE Human colorectal cancer antigen coding sequence SEQ ID NO: 171.

KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.

XX Homo sapiens.

PN WO200155350-A1.

PD 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01350.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457727/49.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
XX disorders related to the colon and rectum including colorectal cancers
XX and also for testing and detection e.g. diagnosis -
XX
XX Disclosure; SEQ ID NO: 171; 522pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of colorectal cancer antigens. These are shown in
XX AA157547-AA157619 and AAM38569-AAM38641. These can be used in the
XX diagnosis, prevention and treatment of cancer of the colon and/or rectum.
XX The present sequence is a colorectal cancer antigen genomic sequence.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 9417 BP; 3043 A; 1599 C; 1757 G; 3012 T; 6 other;
SQ

Query Match 6.9%; Score 19; DB 22; Length 9417;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TCAGTTGGTCAGCCTCCA 24
|||||
Db 4701 TCAGTTGGTCAGCCTCCA 4719

Search completed: January 21, 2003, 13:49:57
Job time : 164.669 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 21, 2003, 10:00:41 ; Search time 1269.45 Seconds
(without alignments)
6877.663 Million cell updates/sec

Title: US-10-060-830-4
Perfect score: 300
Sequence: 1 aaacagaagaaaaaaactg.....tatgcagacccatgaccccta 300

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenBank

- 1: gb_ba.*
- 2: gb_mtg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	300	100.0	2939	9	AF387547	AF387547 Homo sapi
2	300	100.0	6093	9	AB073146	AB073146 Homo sapi
3	195	65.0	3858	9	BC029658	BC029658 Homo sapi
4	140	46.7	192247	2	AC013497	AC013497 Homo sapi
5	140	46.7	193623	9	AC106728	AC106728 Homo sapi
6	109	36.3	146515	2	AC025661	AC025661 Homo sapi
7	58	19.3	217	6	AX182030	AX182030 Sequence
8	58	19.3	1907	6	AX405795	AX405795 Sequence
9	30	10.0	2642	10	AF387548	AF387548 Mus muscu
10	23	7.7	105960	9	HS209A6	AL035401 Human DNA
11	21	7.0	103666	2	AC112542	AC112542 Rattus no
12	21	7.0	160070	2	AC128988	AC128988 Rattus no
13	21	7.0	177157	2	AC094869	AC094869 Rattus no
14	21	7.0	180560	2	AC096787	AC096787 Rattus no
15	21	7.0	180989	2	AC094439	AC094439 Rattus no
16	20	6.7	740	11	G58519	G58519 SHGC-104749
17	20	6.7	8426	2	AC013966	AC013966 Drosophill
18	20	6.7	24492	3	DMC5667	AL031028 Drosophill
19	20	6.7	40047	7	AF440571	AF440571 Sulfolobu
20	20	6.7	106576	2	AC011173	AC011173 Homo sapi
21	20	6.7	107319	2	AC099179	AC099179 Rattus no
22	20	6.7	134156	9	AL357083	AL357083 Human DNA
23	20	6.7	141170	9	AL353607	AL353607 Human DNA
24	20	6.7	148290	9	AC107979	AC107979 Homo sapi
25	20	6.7	150785	2	AC026158	AC026158 Homo sapi
26	20	6.7	153304	9	AC084834	AC084834 Homo sapi
27	20	6.7	154243	2	AC055768	AC055768 Homo sapi
28	20	6.7	154459	2	AL357653	AL357653 Homo sapi
29	20	6.7	160766	9	AC022679	AC022679 Homo sapi
30	20	6.7	161273	2	AC026159	AC026159 Homo sapi
31	20	6.7	163732	2	AC115395	AC115395 Rattus no
32	20	6.7	165038	2	AC027401	AC027401 Homo sapi
33	20	6.7	165138	2	AC068955	AC068955 Homo sapi
34	20	6.7	165649	9	AC103996	AC103996 Homo sapi
35	20	6.7	169355	2	AC114663	AC114663 Mus muscu
36	20	6.7	172261	2	AC111840	AC111840 Rattus no
37	20	6.7	177096	3	AC107326	AC107326 Drosophill
38	20	6.7	179413	2	AC023486	AC023486 Homo sapi
39	20	6.7	182508	2	AC126821	AC126821 Rattus no
40	20	6.7	183384	9	AL669843	AL669843 Human DNA
41	20	6.7	187529	10	AL627131	AL627131 Mouse DNA
42	20	6.7	201465	2	AC122424	AC122424 Mus muscu
43	20	6.7	202984	2	AC097203	AC097203 Rattus no
44	20	6.7	203888	2	AC117148	AC117148 Rattus no
45	20	6.7	206803	2	AC115932	AC115932 Mus muscu

ALIGNMENTS

RESULT 1
AF387547
LOCUS
DEFINITION
2939 bp mRNA linear PRI 11-NOV-2001
Homo sapiens endothelial and smooth muscle cell-derived
neuropilin-like protein (ESDN) mRNA, complete cds.
ACCESSION
AF387547
VERSION
AF387547.1 GI:16902434
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2939)
Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N.,
Matsumori,A., Sasayama,S., Honjo,T. and Tashiro,K.

TITLE ESDN, a novel neuropilin-like membrane protein cloned from vascular cells with the longest secretory signal sequence among eukaryotes, is up-regulated after vascular injury
J. Biol. Chem. 276 (36), 34105-34114 (2001)
MEDLINE 21424027
PUBMED 11447234
REFERENCE 2 (bases 1 to 2939)
AUTHORS Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N., Sasayama, S., Matsumori, A., Honjo, T. and Tashiro, K.
TITLE Direct Submission
JOURNAL Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

FEATURES
source
1..2939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
1..2939
/gene="ESDN"
12..2339
/gene="ESDN"
/codon_start=1
/product="endothelial and smooth muscle cell-derived neuropilin-like protein"
/protein_id="AAL30178.1"
/db_xref="GI:16902435"
/translation="MASRAVVRARRCPQCPQVRAAAXAAPAAALPLSRSLPPCSNSS FSPMLFLLLLVLLLEDAGAQDGGHTVLGPGSGTLTSINYPOTYPSNVCSWE IRVKGMRVRIKFGDFIEDSDCHFNLYRIGVSRTEIGKYGCLGLQMHNSIES KGNETILLFMSGIHVSGRGLASVSDKODLTCLDTASNFLPEPFSKCPAGCLLP FAETISGTHPGYRDSPLCMAGVHAGVSVNTLGGQISVWISKGIPIYESSLANVTSV VGHLSLFTFKSGCYGLMGESGVIAADPOITASSVLENTDHTQENSWKPKARLK KPGPPWAFAIDYQWQIDLNKEKITGTTGTSVNEHNYVSAIRLYSDGQKW TVYREPVEQDKIFQGNKYHQDVNRNPLPIIARFIRVNPTQWQKIAMKMLLGCQ FIPKGRPKLTQPPPRNSNDLKNTPAPKIAKRAKFTQPLQPRSSNEFFPAQTQET TASPDLRNTTVTPNVTKDVALAALVPLVLMVLTLLILVCAWHNRNKKTEGYD LPVGDVTLHORSTFKPEEGKEAGYADLPYNSGQEVHYHAYAEPLPTGPETATPII MDXSGHPTTSVQGFSTFKATGNQPPPLVGTNTLLSRDSCSSAQAQYDTPKAGRP GLPAPDELVYQVQSTQEVSGRDCGVFEIL"
BASE COUNT 825 a 675 c 657 g 782 t

Query Match 100.0%; Score 300; DB 9; Length 2939;
Best Local Similarity 100.0%; Pred. No. 2.6e-159;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAACAGAAGAAAACCTGAAGCACCCTATCACTTACTTCTGAGCAGGCGAGGTG 60
|||||
DB 1676 AAACAGAAGAAAACCTGAAGCACCCTATCACTTACTTCTGAGCAGGCGAGGTG 1735
QY 61 GTGGAAGAGATGAGCAGTTCTTCTGCAAGAGCAGTGGACCATGAGGAACCCAGT 120
|||||
DB 1736 GTGGAAGAGATGAGCAGTTCTTCTGCAAGAGCAGTGGACCATGAGGAACCCAGT 1795
QY 121 TCGCTATAGCAGCAGGAGTTATCACTGAGTCCAGAGAGTCAACAGTGCCTGCA 180
|||||
DB 1796 TCGCTATAGCAGCAGGAGTTATCACTGAGTCCAGAGAGTCAACAGTGCCTGCA 1855
QY 181 GCGTACTCTGCAGATGCTGACGCCACTGTTAGGAGGAATGTTGTACACTTCATCA 240
|||||
DB 1856 GCGTACTCTGCAGATGCTGACGCCACTGTTAGGAGGAATGTTGTACACTTCATCA 1915
QY 241 AGATCTACCTTTAAACAGAGAGGAAAAGAGAGCGGCTATGACACCTAGATCCTTA 300
|||||
DB 1916 AGATCTACCTTTAAACAGAGAGGAAAAGAGAGCGGCTATGACACCTAGATCCTTA 1975
RESULT 2
AB073146 6093 bp mRNA linear PRI 26-APR-2002
LOCUS AB073146
DEFINITION Homo sapiens mRNA for CLCP1, complete cds.

AB073146
AB073146.1 GI:20330503
Homo sapiens lung cell_line:normal lung tissue cDNA to mRNA.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Koshikawa, K., Osada, H., Kozaki, K., Konishi, H., Masuda, A., Tatematsu, Y., Mitsudomi, T., Nakao, A. and Takahashi, T.
Significant up-regulation of a novel gene, CLCP1, in a highly metastatic lung cancer subline as well as in lung cancers in vivo
Oncogene (2002) In press
2 (bases 1 to 6093)
Koshikawa, K., Konishi, H., Osada, H. and Takahashi, T.
Direct Submission
Submitted (16-OCT-2001) Takashi Takahashi, Alchi Cancer Center Research Institute, Division of Molecular Oncology; Chikusa-ku, Kanokoden 1-1, Nagoya, Alchi 464-8681, Japan
(E-mail: takachi@cc.jp, Tel: 81-52-764-2993, Fax: 81-52-764-2993)
Location/Qualifiers
1..6093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3q12.1-12.2"
/cell_line="normal lung tissue"
/tissue_type="lung"
1..6093
/gene="CLCP1"
/gene="CLCP1"
/function="associated with metastatic phenotype"
domains protein 1
/codon_start=1
/evidence=experimental
/product="CLCP1"
/protein_id="BA91138.1"
/db_xref="GI:20330504"
/translation="MASRAVVRARRCPQCPQVRAAAXAAPAAALPLSRSLPPCSNSS FSPMLFLLLLVLLLEDAGAQDGGHTVLGPGSGTLTSINYPOTYPSNVCSWE IRVKGMRVRIKFGDFIEDSDCHFNLYRIGVSRTEIGKYGCLGLQMHNSIES KGNETILLFMSGIHVSGRGLASVSDKODLTCLDTASNFLPEPFSKCPAGCLLP FAETISGTHPGYRDSPLCMAGVHAGVSVNTLGGQISVWISKGIPIYESSLANVTSV VGHLSLFTFKSGCYGLMGESGVIAADPOITASSVLENTDHTQENSWKPKARLK KPGPPWAFAIDYQWQIDLNKEKITGTTGTSVNEHNYVSAIRLYSDGQKW TVYREPVEQDKIFQGNKYHQDVNRNPLPIIARFIRVNPTQWQKIAMKMLLGCQ FIPKGRPKLTQPPPRNSNDLKNTPAPKIAKRAKFTQPLQPRSSNEFFPAQTQET TASPDLRNTTVTPNVTKDVALAALVPLVLMVLTLLILVCAWHNRNKKTEGYD LPVGDVTLHORSTFKPEEGKEAGYADLPYNSGQEVHYHAYAEPLPTGPETATPII MDXSGHPTTSVQGFSTFKATGNQPPPLVGTNTLLSRDSCSSAQAQYDTPKAGRP GLPAPDELVYQVQSTQEVSGRDCGVFEIL"
577..924
/gene="CLCP1"
/note="CUB domain"
934..1191
/gene="CLCP1"
/note="LCCL domain"
1234..1710
/gene="CLCP1"
/note="F5/8 type C domain"
1942..2010
/gene="CLCP1"
/note="transmembrane domain"
4264..4269
/gene="CLCP1"
5873..5878
/gene="CLCP1"
6052..6057
/gene="CLCP1"
BASE COUNT 1671 a 1258 c 1289 g 1874 t 1 others

ORIGIN

Query Match 100.0%; Score 300; DB 9; Length 6093;
 Best Local Similarity 100.0%; Pred. No. 2.5e-159; Indels 0; Gaps 0;
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACAGAAAGAAAACTGAAGCACCTATGACTTACTTACTGGACCGGCGAGTTG 60
 |||||
 Db 2028 AAACAGAAAGAAAACTGAAGCACCTATGACTTACTTACTGGACCGGCGAGTTG 2087

QY 61 GTGGAAAGGAATGAGCAGTTCTTCTCGCAAAAGCAGTGGACCATGAGAAACCCCACT 120
 |||||
 Db 2088 GTGGAAAGGAATGAGCAGTTCTTCTCGCAAAAGCAGTGGACCATGAGAAACCCCACT 2147

QY 121 TCGCTATAGCAGCAGCGAAGTTAATCACTGAGTCCAGAGAGAGTCACACAGTGTGCTGA 180
 |||||
 Db 2148 TCGCTATAGCAGCAGCGAAGTTAATCACTGAGTCCAGAGAGAGTCACACAGTGTGCTGA 2207

QY 181 GGCTGACTCTGCAGAGTATGCTCAGCCACTGTTAGGAGGAATGTTGTGACACTTCAATCA 240
 |||||
 Db 2208 GGCTGACTCTGCAGAGTATGCTCAGCCACTGTTAGGAGGAATGTTGTGACACTTCAATCA 2267

QY 241 AGATCTACCTTTAAACAGAGAGAGAAAGCAGGCTATGACAGACCTAGATCCCTTA 300
 |||||
 Db 2268 AGATCTACCTTTAAACAGAGAGAGAAAGCAGGCTATGACAGACCTAGATCCCTTA 2327

RESULT 3

BC029658 3858 bp mRNA linear PRI 20-MAY-2002
 LOCUS
 DEFINITION Homo sapiens, Similar to endothelial and smooth muscle cell-derived neuropilin-like protein, clone MGC:30023 IMAGE:4431275, mRNA, complete cds.

ACCESSION BC029658
 VERSION BC029658.1 GI:20988614
 KEYWORDS MGC.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 3858)
 Strausberg, R.
 Direct Submission
 Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/Dnp
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAP Plate: 42 Row: c Column: 3.
 Location/Qualifiers

FEATURES

source

1. 3858
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:30023 IMAGE:4431275"
 /tissue_type="Prostate, adenocarcinoma."
 /clone_lib="NTH_MGC_91"
 /lab_host="DH10B"

CDS

/note="Vector: pCMV-SPORT6"
 21..2252
 /codon_start=1
 /product="Similar to endothelial and smooth muscle cell-derived neuropilin-like protein"
 /protein_id="AAH29658.1"
 /db_xref="GI:20988615"
 /translation="MPLFLLLLVLLLLLEDAGAQDGGCHTVLGPESCTLTISINYP
 OTYPSNVCEWEIRVKMGVRIRKFGDFDIEDSDSCHFNLYRNGVGVSTEGKYC
 GLQGMNHSIEKSGNETLLFMSGIHVSGRFLASYSVIDKODLITCLDPAENLEPE
 FSKYCPAGCLLPFAEISGTIPGHVDSPLCMAGVAGVSNLTGGGQISVISKGIPI
 YESLIANNVTSVGHLSLSTLFKTSYSGYGLTGMESGVADIPQITASVLEWHTDTPY
 YRIYSDGQKWTYVREFVQDKIFQGNKDYHQDVRRNPLPPIAIFRNVNPTWQQ
 KIAMKMLLQCOFIPKGRPKLTPPPRPNNDLNTAPPKIAKGRAPKTPQLQPR
 SSNEPQTEQTTASPDINTTPTNTKDALVALVPLVLMVLTLLTLLLVCAWHV
 RNRKKTGTYDLYWDRAAGFYLMVSLACRHNKGWKGKQFLPAKAVDHEETPVYS
 SSEVNLHSPREVTYVLAQDSAEYAPLVGGIVGLTHQSTFKPEEGKEKGVADLDPIN
 SPQGEVYHAYAEPLPTITGPEYATPIIMDSGHPTTSVGQPSSTFKATGNPPPLVGT
 YNTLLSRITDSCSAQAQYDTPKAGPLPAPDELVYQVPSQTEVSGAGRGDCDVEK
 EIL"

BASE COUNT 1137 a 807 c 785 g 1129 t
 ORIGIN

Query Match 65.0%; Score 195; DB 9; Length 3858;

Best Local Similarity 99.6%; Pred. No. 1.6e-99;

Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 55 AGTTGGTGGAAAGAAATGAACAGCTTCTTCTGCAAAAGCAGTGGACCATGAGGAAC 114

|||||

Db 1643 AGTTGGTGGAAAGAAATGAACAGCTTCTTCTGCAAAAGCAGTGGACCATGAGGAAC 1702

|||||

QY 115 CCAGTTCCTATAGCAGCAGCAAGTTAATCACCTGAGTCCCAAGAGAGTCACACAGT 174

|||||

Db 1703 CCAGTTCCTATAGCAGCAGTGAAGTTAATCACCTGAGTCCCAAGAGAGTCACACAGT 1762

|||||

QY 175 GCTGAGGCTGACTCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTACACT 234

|||||

Db 1763 GCTGAGGCTGACTCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTACACT 1822

|||||

QY 235 TCATCAAGATCTACCTTTAAACAGAGAGAAAGAGAGGAGCTATGCGAGACCTAGA 294

|||||

Db 1823 TCATCAAGATCTACCTTTAAACAGAGAGAAAGAGAGGAGCTATGCGAGACCTAGA 1882

|||||

QY 295 TCCTTA 300

Db 1883 TCCTTA 1888

RESULT 4

AC013497 192247 bp DNA linear HTG 26-MAR-2000
 LOCUS
 DEFINITION Homo sapiens chromosome 3 clone RP11-319J24 map 3, WORKING DRAFT SEQUENCE, 11 unordered pieces.

ACCESSION AC013497
 VERSION AC013497.4 GI:7329299
 KEYWORDS HTG; HTGS-PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 192247)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 3, clone RP11-319J24
 Unpublished
 2 (bases 1 to 192247)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Brown, A., Castle, A., Colangelo, M., Collins, S., Collins, S., Collman, A., Cooke, P., Dearrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 1 (bases 1 to 192247)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Brown, A., Castle, A., Colangelo, M., Collins, S., Collins, S., Collman, A., Cooke, P., Dearrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,

Lehoczyk, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6692347.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: LJ852
Center clone name: 319_J_24
----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182870 bases at least Q40
Consensus quality: 187032 bases at least Q30
Consensus quality: 188925 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 191247; sum-of-contigs
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-contigs

***** NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 472: contig of 472 bp in length
* 473 572: gap of 100 bp
* 573 3945: contig of 3373 bp in length
* 3946 4045: gap of 100 bp
* 4046 11056: contig of 7011 bp in length
* 11057 11156: gap of 100 bp
* 11157 17728: contig of 6572 bp in length
* 17729 17828: gap of 100 bp
* 17829 30556: contig of 12728 bp in length
* 30557 30556: gap of 100 bp
* 30557 49294: contig of 18638 bp in length
* 49295 49394: gap of 100 bp
* 49395 64509: contig of 15115 bp in length
* 64510 64609: gap of 100 bp
* 64610 83042: contig of 18433 bp in length
* 83043 83142: gap of 100 bp
* 83143 112373: contig of 29231 bp in length
* 112374 112473: gap of 100 bp
* 112474 145663: contig of 33190 bp in length
* 145664 145763: gap of 100 bp
* 145764 192247: contig of 46484 bp in length.

FEATURES source

1. .192247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-319J24"
/clone_lib="RPCI-11 Human Male BAC"
1. 472
/note="assembly_fragment"

misc_feature

clone_end:SP6
vector_side:right
573. .3945
/note="assembly_fragment"
4046. .11056
/note="assembly_fragment"
11157. .17728
/note="assembly_fragment"
17829. .30556
/note="assembly_fragment"
30657. .49294
/note="assembly_fragment"
49395. .64509
/note="assembly_fragment"
64610. .83042
/note="assembly_fragment"
clone_end:T7
vector_side:left
83143. .112373
/note="assembly_fragment"
112474. .145663
/note="assembly_fragment"
145764. .192247
/note="assembly_fragment"
BASE COUNT 58524 a 38157 c 36840 g 57721 t 1005 others
ORIGIN
Query Match 46.7%; Score 140; DB 2; Length 192247;
Best Local Similarity 100.0%; Pred. No. 2.8e-68;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 55 AGTTGGTGGGAAGGAAATGAAGCAGTTTCTTCTGCAAGAGAGTGCACATGAGGAAC 114
DB 164101 AGTTGGTGGGAAGGAAATGAAGCAGTTTCTTCTGCAAGAGAGTGCACATGAGGAAC 164160
QY 115 CCACGTCCTATAGCAGCAGGAGTAAATCAGCTGAGTCCAAAGAGAGTGCACACAGT 174
DB 164161 CCACGTCCTATAGCAGCAGGAGTAAATCAGCTGAGTCCAAAGAGAGTGCACACAGT 164220
QY 175 GCTGCAGGCTGACTCTGCAG 194
DB 164221 GCTGCAGGCTGACTCTGCAG 164240
RESULT 5
AC106728 193623 bp DNA linear PRI 25-MAY-2002
LOCUS Homo sapiens 3 BAC RP11-319J24 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC106728
VERSION AC106728.6 GI:21205081
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 193623)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbala, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X.,


```

repeat_region complement(19662..19780)
repeat_region /rpt_family="FLAM_A"
repeat_region complement(19781..21049)
repeat_region /rpt_family="L1MD3"
repeat_region complement(21069..21455)
repeat_region /rpt_family="MT1B"
repeat_region complement(21479..22318)
repeat_region /rpt_family="L1MD3"
repeat_region complement(22447..22681)
repeat_region /rpt_family="L1MD3"
repeat_region complement(23470..23853)
repeat_region /rpt_family="MSTD"
repeat_region 23852..23925
repeat_region /rpt_family="U3"
repeat_region 23936..25203
repeat_region /rpt_family="L1PB3"
STS 24844..24975
/standard_name="6234"
repeat_region 25219..25388
repeat_region /rpt_family="TTTCn"
repeat_region complement(25428..25548)
repeat_region /rpt_family="AluJb"
repeat_region 25555..25950
repeat_region /rpt_family="L1PB3"
repeat_region 25951..26456
repeat_region /rpt_family="MER75"
repeat_region 26457..27565

```

```

Query Match 46.7% Score 140; DB 9; Length 193623;
Best Local Similarity 100.0%; Pred. No. 2.8e-68;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 AGTTGGTGAAGGAATGAACGAGTTCTCTCTCAAAAGCAGTGGACCATGAGGAAC 114
Db 95314 AGTTGGTGAAGGATGAACGAGTTCTCTCTCAAAAGCAGTGGACCATGAGGAAC 95373

QY 115 CCAGTTCCTATAGCAGCAGCAAGTAAATCACTGATCCAGAGAGTCACCAACAGT 174
Db 95374 CCAGTTCCTATAGCAGCAGCAAGTAAATCACTGATCCAGAGAGTCACCAACAGT 95433

QY 175 GCTGAGGCTGACTCTGCAG 194
Db 95434 GCTGAGGCTGACTCTGCAG 95453

```

```

RESULT 6
AC025661
LOCUS AC025661 Homo sapiens chromosome 3 clone RP11-322A20 map 3, LOW-PASS
DEFINITION AC025661 Homo sapiens chromosome 3 clone RP11-322A20 map 3, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC025661
VERSION AC025661.2 GI:7543781
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 146515)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castie,A., Choepel,Y., Colangelo,M., Collins,S.,
Collumore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Unpublished
Homo sapiens chromosome 3, clone RP11-322A20
2 (bases 1 to 146515)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castie,A., Choepel,Y., Colangelo,M., Collins,S.,
Collumore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

```

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 146515)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campoliano,A., Castie,A., Choepel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 12, 2000 this sequence version replaced gi:7230271.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7942

Center clone name: 322_A_20

* NOTE: This record contains 164 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

* 1 836: contig of 836 bp in length

* 837 936: gap of 100 bp

* 937 1748: contig of 812 bp in length

* 1749 1848: gap of 100 bp

* 1849 2680: contig of 832 bp in length

* 2681 2780: gap of 100 bp

* 2781 3534: contig of 754 bp in length

* 3535 3634: gap of 100 bp

* 3635 4383: contig of 749 bp in length

* 4384 4483: gap of 100 bp

* 4484 5261: contig of 778 bp in length

* 5261: contig of 778 bp in length

* 5261: contig of 778 bp in length

* 5261: contig of 778 bp in length

* 5261: contig of 778 bp in length

* 5261: contig of 778 bp in length

* 5261: contig of 778 bp in length

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known, annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, ENBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/projects/Celegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

ncsp://www.bacpac.med.buffalo.edu/

VECTION: PERFECT
This sequence is the entire insert of clone RP1-209A6.

FEATURES

source

URES	source	Location/Qualifiers
		1. .105960
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="6"
		/map="p22.1-24.3"
		/clone="RPI-209A6"
		/clone_lib="RPCI-1"
		2801. .3175
misc_feature		/note="match: GSS: Em:AQ520434"
repeat_region		4571. .4616
repeat_region		/note="MER91C repeat: matches 37. .84 of consensus"
repeat_region		4630. .4721
repeat_region		/note=".46 copies 2 mer ta 90% conserved"
repeat_region		5464. .5812
repeat_region		/note="MSTA repeat: matches 1. .409 of consensus"
misc_feature		5924. .6327
misc_feature		/note="MLT1B repeat: matches 1. .390 of consensus"
misc_feature		complement(6546. .7247)
misc_feature		/note="match: GSS: Em:AQ532066"
misc_feature		complement(6613. .7267)
misc_feature		/note="match: GSS: Em:AQ390074"
misc_feature		complement(6762. .7231)
misc_feature		/note="match: GSS: Em:AQ720342"
misc_feature		complement(6845. .7266)
repeat_region		/note="match: GSS: Em:AQ618414"
repeat_region		7936. .8121
repeat_region		/note="MERSA repeat: matches 1. .189 of consensus"
misc_feature		9455. .9851
repeat_region		/note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region		complement(9923. .10429)
repeat_region		/note="match: GSS: Em:AQ519422"
repeat_region		11595. .11688
repeat_region		/note=".47 copies 2 mer aa 62% conserved"
repeat_region		11911. .12215
misc_feature		/note="AluSp repeat: matches 1. .313 of consensus"
misc_feature		12241. .12910
misc_feature		/note="match: GSS: Em:AQ423435"
misc_feature		12247. .12726
misc_feature		/note="match: GSS: Em:AQ209760"
repeat_region		12345. .12910
repeat_region		/note="match: GSS: Em:AQ428036"
repeat_region		13474. .15062
repeat_region		/note="trigger4(Zombi) repeat: matches 1. .1616 of consensus"
repeat_region		15063. .15334
repeat_region		/note="AluX repeat: matches 29. .299 of consensus"
repeat_region		15335. .16456
repeat_region		/note="trigger4(Zombi) repeat: matches 1616. .2729 of consensus"
repeat_region		16505. .16546
repeat_region		/note=".21 copies 2 mer at 76% conserved"
repeat_region		16650. .16905

17147. .17623
/note=LiM4 repeat: matches 47597. .5019 of consensus"
/note=LiM6 repeat: matches 5261. .5488 of consensus"
18793. .18936
/note=72 copies 2 mer tt 57% conserved"
19014. .19212
/note=Li2 repeat: matches 1924. .2140 of consensus"
19255. .19566
/note=AluY repeat: matches 1. .310 of consensus"
20588. .20811
/note=LiM4 repeat: matches 4487. .4715 of consensus"
20812. .21137
/note=LiM12 repeat: matches -1416. -.1097 of consensus"
20954. .21529
/note="match: GSS: Em:AQ237509"
complement(20980. .21565)
/note="match: GSS: Em:AQ380004"
complement(21036. .21560)
/note="match: GSS: Em:AQ486517"
21431. .24498
/note=LiP3 repeat: matches -434. .3412 of consensus"
24515. .24563
/note=LiP repeat: matches' 4408. .4456 of consensus"
24562. .25046
/note=LiP repeat: matches 3570. .4056 of consensus"
25048. .25356
/note=LiP7 repeat: matches 5803. .6134 of consensus"
25362. .25383
/note="11 copies 2 mer ca 100% conserved"
25846. .25955
/note=LiM4 repeat: matches 5199. .5320 of consensus"
25956. .26255
/note=AluSx repeat: matches 1. .296 of consensus"
26256. .26319
/note=LiM4 repeat: matches 5320. .5380 of consensus"
26320. .27507
/note=LiM6 repeat: matches 2913. .3803 of consensus"
27508. .27639
/note=FLAM A repeat: matches 1. .123 of consensus"
27640. .28203
/note=LiM6C repeat: matches 2363. .2913 of consensus"
28297. .28954
/note=LiP7 repeat: matches 5485. .6143 of consensus"
28955. .29255
/note=AluY repeat: matches 1. .311 of consensus"
29256. .32758
/note=LiP7 repeat: matches 1959. .5485 of consensus"
33341. .34052
/note=LiM2 repeat: matches 974. .1732 of consensus"
35064. .35178
/note=FLAM C repeat: matches 2. .127 of consensus"
35475. .35698
/note=WR repeat: matches 7. .248 of consensus"
35857. .36206
/note=LiM6A repeat: matches 5963. .6300 of consensus"
36833. .37388
/note="match: STS: Em:Z78722"
37524. .37573
/note="25 copies 2 mer aa 72% conserved"
38019. .38074
/note=Li2 repeat: matches 2653. .2705 of consensus"
40368. .41178
/note=LiM1 repeat: matches 5318. .6163 of consensus"
41193. .41970
/note=LiP14 repeat: matches 4775. .5553 of consensus"
41971. .42381
/note=LiP14 repeat: matches 5734. .6149 of consensus"
42445. .42665
/note=LiM4 repeat: matches 5061. .5273 of consensus"
43158. .43315
/note=MERSA repeat: matches 7. .159 of consensus"
45347. .45448
/note=MERSA repeat: matches 1. .104 of consensus"

```

misc_feature      complement(45414..45971)
misc_feature      /note="match: GSS: Em:B52631"
misc_feature      complement(45526..45807)
misc_feature      /note="match: STS: Em:G54793"
repeat_region     45977..46264 GSS: Em:B30942
misc_feature      /note="match: GSS: Em:B30942"
repeat_region     48380..48531
misc_feature      /note="76 copies 2 mer at 66% conserved"
repeat_region     complement(48635..49059)
misc_feature      /note="match: GSS: Em:AQ757741"
repeat_region     48835..49217
repeat_region     /note="THEIC repeat: matches 15..369 of consensus"
misc_feature      49235..49287
repeat_region     /note="L2 repeat: matches 2591..2642 of consensus"
misc_feature      49378..49854
repeat_region     /note="match: GSS: Em:AQ621055"
misc_feature      49389..49608
repeat_region     /note="MER30 repeat: matches 1..230 of consensus"
misc_feature      49397..49845
repeat_region     /note="match: GSS: Em:AQ117725"
misc_feature      50308..50706
repeat_region     /note="MLR2CB repeat: matches 1..462 of consensus"
repeat_region     50711..50758
repeat_region     /note="74 copies 2 mer to 87% conserved"
repeat_region     51298..51356
repeat_region     /note="MADE1 repeat: matches 22..80 of consensus"
repeat_region     51773..52088
repeat_region     /note="L2 repeat: matches 1667..2013 of consensus"
repeat_region     52439..52791
repeat_region     /note="LTR40B repeat: matches 1..375 of consensus"
repeat_region     52792..53231

Query Match      7.7%; Score 23; DB 9; Length 105960;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 AGCACTGGACCATGAGGAACCC 116
|||||
Db 45795 AGCACTGGACCATGAGGAACCC 45817

RESULT 11
AC112542
LOCUS
DEFINITION      Rattus norvegicus clone CH230-10C21, *** SEQUENCING IN PROGRESS
ACCESSION      AC112542
VERSION        AC112542.2 GI:21731372
KEYWORDS       HTG; HTGS_PHASE1.
SOURCE         Norway rat.
ORGANISM       Rattus norvegicus
Eukaryota; Eutheria; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1. (bases 1 to 103666)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leai,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Mitchell,T., Mohabbat,K., Morgan,M., Morris,N.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Reay,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanik,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Unpublished
Direct Submission
2 (bases 1 to 103666)
Worley,K.C.
Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 103666)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:18860128.
----- Genome Center of Medicine
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTW
Center clone name: CH230-10C21
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 59568 bases at least Q40
Consensus quality: 62320 bases at least Q30
Consensus quality: 64819 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1170: contig of 1170 bp in length
* 1270: gap of unknown length
* 1271: contig of 1583 bp in length
* 2853: gap of unknown length
* 2854: contig of 1506 bp in length
* 2954: gap of unknown length
* 4450: gap of unknown length
* 4553: gap of unknown length
* 6200: contig of 1641 bp in length
* 6301: gap of unknown length
* 7482: contig of 1182 bp in length
* 7483: gap of unknown length

```

* 7593 contig of 1241 bp in length
 * 8823: gap of unknown length
 * 8923: contig of 1340 bp in length
 * 10263: gap of unknown length
 * 10363: gap of unknown length
 * 11442: contig of 1079 bp in length
 * 11542: gap of unknown length
 * 11543: contig of 1064 bp in length
 * 12607: gap of unknown length
 * 12707: contig of 1030 bp in length
 * 13737: gap of unknown length
 * 13837: contig of 1479 bp in length
 * 15315: gap of unknown length
 * 15316: contig of 1072 bp in length
 * 16487: gap of unknown length
 * 16488: contig of 1227 bp in length
 * 16588: gap of unknown length
 * 17914: gap of unknown length
 * 17915: contig of 1540 bp in length
 * 19455: gap of unknown length
 * 19455: contig of 1141 bp in length
 * 20695: gap of unknown length
 * 20696: contig of 1125 bp in length
 * 20796: gap of unknown length
 * 21921: contig of 1324 bp in length
 * 22021: gap of unknown length
 * 23344: gap of unknown length
 * 23445: contig of 1457 bp in length
 * 23451: gap of unknown length
 * 24902: contig of 1570 bp in length
 * 25002: gap of unknown length
 * 26572: contig of 1313 bp in length
 * 26672: gap of unknown length
 * 27985: contig of 1645 bp in length
 * 28085: gap of unknown length
 * 28729: gap of unknown length
 * 29730: contig of 2186 bp in length
 * 29830: gap of unknown length
 * 32115: gap of unknown length
 * 32116: contig of 1711 bp in length
 * 33827: gap of unknown length
 * 33927: contig of 1597 bp in length
 * 35524: gap of unknown length
 * 35624: contig of 1796 bp in length
 * 37420: gap of unknown length
 * 37520: contig of 1611 bp in length
 * 39131: gap of unknown length
 * 39231: contig of 1887 bp in length
 * 41117: gap of unknown length
 * 41218: contig of 1709 bp in length
 * 42927: gap of unknown length
 * 43027: contig of 1992 bp in length
 * 45019: gap of unknown length
 * 45119: contig of 1405 bp in length
 * 46524: gap of unknown length
 * 46624: contig of 3280 bp in length
 * 49904: gap of unknown length
 * 50003: contig of 2692 bp in length
 * 52695: gap of unknown length
 * 52696: contig of 2269 bp in length
 * 52796: gap of unknown length
 * 53064: contig of 1943 bp in length
 * 55164: gap of unknown length
 * 55165: contig of 1943 bp in length
 * 57108: gap of unknown length
 * 57208: contig of 1301 bp in length
 * 58509: gap of unknown length
 * 58609: contig of 1571 bp in length
 * 60180: gap of unknown length
 * 60279: contig of 2579 bp in length
 * 62858: gap of unknown length
 * 62959: contig of 1943 bp in length
 * 64902: gap of unknown length
 * 65001: contig of 2300 bp in length
 * 67301: gap of unknown length
 * 67402: contig of 1685 bp in length
 * 69087: gap of unknown length
 * 69187: contig of 2244 bp in length
 * 71430: gap of unknown length
 * 71530: contig of 4371 bp in length
 * 75901: contig of 4371 bp in length

* 75902 76001: gap of unknown length
 * 76002 78764: contig of 2763 bp in length
 * 78765 78864: gap of unknown length
 * 78865 82421: contig of 3557 bp in length
 * 82422 85613: gap of unknown length
 * 85614 85713: gap of unknown length
 * 85714 85778: contig of 4065 bp in length
 * 89779 89879: gap of unknown length
 * 89879 92607: contig of 2729 bp in length
 * 92608 92708: gap of unknown length
 * 92708 96222: contig of 3515 bp in length
 * 96223 96323: gap of unknown length
 * 96323 103666: contig of 7344 bp in length.
 FEATURES Location/Qualifiers
 source
 1. 103666
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-10C21"
 BASE COUNT 28840 a 20327 c 20378 g 29257 t 4864 others
 ORIGIN
 Query Match 7.0% Score 21; DB 2; Length 103666;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 183 CTGACTCTGCAGATGCTC 203
 ||||||||||||||||||
 Db 61012 CTGACTCTGCAGATGCTC 61032
 RESULT 12
 AC128988 Rattus norvegicus clone CH230-1A18, *** SEQUENCING IN PROGRESS ***,
 LOCUS 160070 bp DNA linear HTG 24-JUL-2002
 DEFINITION 85 unordered pieces.
 AC128988
 AC128988.1 GI:21953882
 VERSION HTG; HTGS_PHASE1.
 KEYWORDS Rattus norvegicus.
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 160070)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbara, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.K., Delgado, O., Denn, A.L., Ding, X., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louise, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metsker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okwuonu, G.,

Oraqunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Roife, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmari, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 160070)
 Worley, K.C.
 Direct Submission
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center of Medicine
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: nUGS
 Center clone name: CH230-1A18
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 75990 bases at least Q40
 Consensus quality: 84091 bases at least Q30
 Consensus quality: 87569 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 85 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1092: contig of 1092 bp in length
 * 1093 1192: gap of unknown length
 * 1193 2207: contig of 1015 bp in length
 * 2208 2307: gap of unknown length
 * 2308 3349: contig of 1042 bp in length
 * 3350 3449: gap of unknown length
 * 3450 4449: contig of 1000 bp in length
 * 4450 4549: gap of unknown length
 * 4550 5571: contig of 1022 bp in length
 * 5572 5671: gap of unknown length
 * 5672 7278: contig of 1607 bp in length
 * 7279 7378: gap of unknown length
 * 7379 8453: contig of 1075 bp in length
 * 8454 8553: gap of unknown length
 * 8554 9601: contig of 1048 bp in length
 * 9602 9702: gap of unknown length
 * 9702 11162: contig of 1461 bp in length
 * 11163 11262: gap of unknown length
 * 11263 12278: contig of 1016 bp in length
 * 12279 12378: gap of unknown length
 * 12379 13379: contig of 1001 bp in length
 * 13380 13479: gap of unknown length
 * 13480 14669: contig of 1190 bp in length
 * 14670 14769: gap of unknown length
 * 14770 15967: contig of 1198 bp in length
 * 15968 16067: gap of unknown length
 * 16068 17268: contig of 1201 bp in length
 *
 * 17269 17368: gap of unknown length
 * 17369 18497: contig of 1129 bp in length
 * 18498 18597: gap of unknown length
 * 18598 19921: contig of 1324 bp in length
 * 19922 20021: gap of unknown length
 * 20022 21668: contig of 1647 bp in length
 * 21669 21768: gap of unknown length
 * 21769 22941: contig of 1172 bp in length
 * 22941 23040: gap of unknown length
 * 23041 24576: contig of 1636 bp in length
 * 24577 24777: gap of unknown length
 * 24778 25940: contig of 1163 bp in length
 * 25941 26040: gap of unknown length
 * 26040 27467: contig of 1428 bp in length
 * 27468 27567: gap of unknown length
 * 27568 29077: contig of 1510 bp in length
 * 29078 29177: gap of unknown length
 * 29178 30187: contig of 1009 bp in length
 * 30188 30287: gap of unknown length
 * 30288 31795: contig of 1508 bp in length
 * 31796 31894: gap of unknown length
 * 31895 33153: contig of 1259 bp in length
 * 33154 33253: gap of unknown length
 * 33254 35176: contig of 1922 bp in length
 * 35177 35276: gap of unknown length
 * 35277 36957: contig of 1682 bp in length
 * 36958 37058: gap of unknown length
 * 37059 38711: contig of 1853 bp in length
 * 38712 38810: gap of unknown length
 * 38811 39843: contig of 1033 bp in length
 * 39844 39943: gap of unknown length
 * 39944 41375: contig of 1431 bp in length
 * 41376 41475: gap of unknown length
 * 41476 42677: contig of 1203 bp in length
 * 42678 42777: gap of unknown length
 * 42778 43919: contig of 1142 bp in length
 * 43920 44019: gap of unknown length
 * 44020 45657: contig of 1638 bp in length
 * 45658 45757: gap of unknown length
 * 45758 47686: contig of 1929 bp in length
 * 47687 47787: gap of unknown length
 * 47788 49082: contig of 1295 bp in length
 * 49083 49182: gap of unknown length
 * 49183 50325: contig of 1144 bp in length
 * 50326 50425: gap of unknown length
 * 50426 51812: contig of 1387 bp in length
 * 51813 51912: gap of unknown length
 * 51913 53186: contig of 1274 bp in length
 * 53187 53286: gap of unknown length
 * 53287 54800: contig of 1514 bp in length
 * 54801 54900: gap of unknown length
 * 54901 56942: contig of 2042 bp in length
 * 56943 57042: gap of unknown length
 * 57043 58593: contig of 1551 bp in length
 * 58594 58693: gap of unknown length
 * 58694 60439: contig of 1745 bp in length
 * 60440 60538: gap of unknown length
 * 60539 61933: contig of 1395 bp in length
 * 61934 62034: gap of unknown length
 * 62035 64041: contig of 2007 bp in length
 * 64042 65154: contig of 1014 bp in length
 * 65155 65254: gap of unknown length
 * 65255 67267: contig of 2012 bp in length
 * 67268 67366: gap of unknown length
 * 67367 69179: contig of 1813 bp in length
 * 69180 69279: gap of unknown length
 * 69280 71708: contig of 2429 bp in length
 * 71709 71808: gap of unknown length
 * 71809 73441: contig of 1633 bp in length
 * 73442 73541: gap of unknown length
 * 73542 76199: contig of 2658 bp in length
 * 76200 76299: gap of unknown length

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

* 76300 78198: contig of 1899 bp in length
 * 78199 78298: gap of unknown length
 * 79982: contig of 1684 bp in length
 * 79983 80083: gap of unknown length
 * 80083 81752: contig of 1670 bp in length
 * 81753 81852: gap of unknown length
 * 81853 83812: contig of 1960 bp in length
 * 83813 83912: gap of unknown length
 * 83913 85574: contig of 1662 bp in length
 * 85575 85674: gap of unknown length
 * 85675 87862: contig of 2188 bp in length
 * 87863 87962: gap of unknown length
 * 87963 90559: contig of 2597 bp in length
 * 90560 90659: gap of unknown length

Query Match 7.0%; Score 21; DB 2; Length 160070;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 58 TTGGTGAAGGAATGAAGCA 78
 |||||
 Db 141766 TTGGTGAAGGAATGAAGCA 141786

RESULT 13
 AC094869/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-5N16, *** SEQUENCING IN PROGRESS ***,
 59 unordered pieces.
 ACCESSION AC094869
 VERSION AC094869.3 (GI:21716223
 KEYWORDS HTG; HTGS-PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 177157)
 Muny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbaria,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H.,
 Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,K., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojubenkan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 177157)
 Worley,K.C.
 Direct Submission
 Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 177157)
 Worley,K.C.
 Direct Submission
 Submitted (09-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 9, 2002 this sequence version replaced gi:17941653.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GBPF
 Center clone name: CH230-5N16
 ----- Summary Statistics

Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 134840 bases at least Q40
 Consensus quality: 142064 bases at least Q30
 Consensus quality: 147026 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 59 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1330: contig of 1330 bp in length
 * 1331 1430: gap of unknown length
 * 1431 2482: contig of 1052 bp in length
 * 2483 2582: gap of unknown length
 * 2583 3702: contig of 1120 bp in length
 * 3703 3802: gap of unknown length
 * 3803 4925: contig of 1123 bp in length
 * 4926 5025: gap of unknown length
 * 5026 6225: contig of 1200 bp in length
 * 6226 6325: gap of unknown length
 * 6326 7610: contig of 1285 bp in length
 * 7611 7710: gap of unknown length
 * 7711 9339: contig of 1529 bp in length
 * 9340 10482: contig of 1143 bp in length
 * 10483 10582: gap of unknown length
 * 10583 12354: contig of 1772 bp in length
 * 12355 12454: gap of unknown length
 * 12455 13585: contig of 1131 bp in length
 * 13586 13685: gap of unknown length
 * 13686 15555: contig of 1870 bp in length
 * 15556 15656: gap of unknown length
 * 15656 18232: contig of 2577 bp in length
 * 18233 18332: gap of unknown length
 * 18333 19689: contig of 1357 bp in length
 * 19690 21501: contig of 1712 bp in length
 * 19790

21502 21601: gap of unknown length
 21602 23691: contig of 2090 bp in length
 23692 23791: gap of unknown length
 23792 26400: contig of 2609 bp in length
 26401 28500: gap of unknown length
 28501 28244: contig of 1744 bp in length
 28245 28344: gap of unknown length
 28345 30087: contig of 1743 bp in length
 30088 30187: gap of unknown length
 30188 32088: contig of 1901 bp in length
 32089 32188: gap of unknown length
 32189 34705: contig of 2517 bp in length
 34706 34805: gap of unknown length
 34806 37570: contig of 2765 bp in length
 37571 37670: gap of unknown length
 37671 40151: contig of 2481 bp in length
 40152 40251: gap of unknown length
 40252 43232: contig of 2981 bp in length
 43233 43332: gap of unknown length
 43333 45819: contig of 2487 bp in length
 45820 45919: gap of unknown length
 45920 48603: contig of 2684 bp in length
 48604 48703: gap of unknown length
 48704 50800: contig of 2097 bp in length
 50801 50900: gap of unknown length
 50901 53660: contig of 2760 bp in length
 53661 53760: gap of unknown length
 53761 55705: contig of 1945 bp in length
 55706 55805: gap of unknown length
 55806 57733: contig of 1928 bp in length
 57734 57833: gap of unknown length
 57834 59881: contig of 2048 bp in length
 59882 59981: gap of unknown length
 59982 63078: contig of 3097 bp in length
 63079 63178: gap of unknown length
 63179 66533: contig of 3355 bp in length
 66534 66633: gap of unknown length
 66634 69341: contig of 2708 bp in length
 69342 69441: gap of unknown length
 69442 72365: contig of 2924 bp in length
 72366 72465: gap of unknown length
 72466 75151: contig of 2686 bp in length
 75152 75251: gap of unknown length
 75252 77337: contig of 2086 bp in length
 77338 77437: gap of unknown length
 77438 80762: contig of 3325 bp in length
 80763 80862: gap of unknown length
 80863 83933: contig of 3071 bp in length
 83934 84033: gap of unknown length
 84034 87443: contig of 3410 bp in length
 87444 87543: gap of unknown length
 87544 90017: contig of 2474 bp in length
 90018 90117: gap of unknown length
 90118 94051: contig of 3934 bp in length
 94052 94151: gap of unknown length
 94152 97609: contig of 3458 bp in length
 97610 97709: gap of unknown length
 100295: contig of 2586 bp in length
 100395: gap of unknown length
 102395: contig of 2344 bp in length
 102740 102839: gap of unknown length
 102840 107839: contig of 5000 bp in length
 107840 107939: gap of unknown length
 111843: contig of 3904 bp in length
 111844 111943: gap of unknown length
 111944 116604: contig of 4661 bp in length
 116605 116704: gap of unknown length
 116705 120324: contig of 3620 bp in length
 120325 120424: gap of unknown length
 120425 123419: contig of 2995 bp in length
 123420 123519: gap of unknown length
 123520 127981: contig of 4462 bp in length
 127982 128081: gap of unknown length

* 128082 132803: contig of 4722 bp in length
 * 132804 132903: gap of unknown length
 * 132904 137151: contig of 4248 bp in length
 * 137152 137251: gap of unknown length
 * 137252 142589: contig of 5338 bp in length
 * 142590 142689: gap of unknown length
 * 142690 146761: contig of 4072 bp in length
 Query Match 7.0%; Score 21; DB 2; Length 177157;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 58 TTGGTGGAAAGGAATGAAGCA 78
 |||||
 Db 61859 TTGGTGGAAAGGAATGAAGCA 61839
 RESULT 14
 AC096787
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-5A1, *** SEQUENCING IN PROGRESS ***
 AC096787
 ACCESSION
 VERSION AC096787.3 GI:21716278
 KEYWORDS HTG; HTGS, PHASE1.
 SOURCE Norway rat
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE
 1 (bases 1 to 180560)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayale, M., Banks, T.,
 Barbara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Briefe, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabriel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Owiedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojebokan, I., Roife, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmari, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.
 Direct Submission
 Unpublished

TITLE
 JOURNAL

REFERENCE

2 (bases 1 to 180560)

Worley, K.C.

Direct Submission

Submitted (28-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 180560)

Worley, K.C.

Direct Submission

Submitted (09-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jul 9, 2002 this sequence version replaced gi:17941378.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GA2X

Center clone name: CH230-5A1

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 120335 bases at least Q40

Consensus quality: 127623 bases at least Q30

Consensus quality: 133514 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 60 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1129: contig of 1129 bp in length

* 1130: gap of unknown length

* 1230: contig of 1620 bp in length

* 2849: gap of unknown length

* 2950: contig of 1610 bp in length

* 4559: gap of unknown length

* 4560: contig of 1379 bp in length

* 6039: gap of unknown length

* 6139: contig of 1045 bp in length

* 7184: gap of unknown length

* 8880: contig of 1597 bp in length

* 8981: gap of unknown length

* 10123: contig of 1142 bp in length

* 10223: gap of unknown length

* 11536: contig of 1314 bp in length

* 11537: gap of unknown length

* 11637: contig of 1424 bp in length

* 13061: gap of unknown length

* 13161: contig of 1138 bp in length

* 14299: gap of unknown length

* 14399: contig of 1410 bp in length

* 15809: gap of unknown length

* 15909: contig of 1452 bp in length

* 17361: gap of unknown length

* 17461: contig of 1075 bp in length

* 18536: gap of unknown length

* 18636: contig of 1795 bp in length

* 20431: gap of unknown length

* 20531: contig of 1197 bp in length

* 21728: gap of unknown length

* 21828: contig of 1524 bp in length

* 23352: gap of unknown length

* 23452: contig of 1563 bp in length

* 25015: gap of unknown length

25115

26547

26647

28476

28576

31006

31106

32691

32791

35315

35415

36603

36703

38656

38756

40271

40371

40372

42040

42140

43801

43901

46568

46668

49303

49403

51419

51519

54228

54328

57002

57102

57594

59694

59695

61551

61651

64380

64480

66969

67069

69609

69709

72000

72100

75154

75254

78190

78290

81877

81977

85416

85516

89542

89642

91907

92007

96062

96162

99760

99860

103562

103562

103662

106541

106641

110563

110663

114867

114967

118736

118836

123825

123925

130817

130917

26546: contig of 1432 bp in length

26646: gap of unknown length

28475: contig of 1829 bp in length

28575: gap of unknown length

31005: contig of 2430 bp in length

31105: gap of unknown length

32690: contig of 1585 bp in length

32790: gap of unknown length

35314: contig of 2524 bp in length

35414: gap of unknown length

36602: contig of 1188 bp in length

36702: gap of unknown length

38655: contig of 1953 bp in length

38755: gap of unknown length

40271: contig of 1516 bp in length

40371: gap of unknown length

42035: contig of 1668 bp in length

42135: gap of unknown length

43800: contig of 1661 bp in length

43900: gap of unknown length

46567: contig of 2667 bp in length

46667: gap of unknown length

49302: contig of 2635 bp in length

49402: gap of unknown length

51418: contig of 2016 bp in length

51518: gap of unknown length

54227: contig of 2709 bp in length

54327: gap of unknown length

57002: contig of 2675 bp in length

57102: gap of unknown length

59594: contig of 2492 bp in length

59694: gap of unknown length

61550: contig of 1856 bp in length

61650: gap of unknown length

64379: contig of 2729 bp in length

64479: gap of unknown length

66968: contig of 2489 bp in length

67068: gap of unknown length

69608: contig of 2540 bp in length

69708: gap of unknown length

71999: contig of 2291 bp in length

72099: gap of unknown length

75153: contig of 3054 bp in length

75253: gap of unknown length

78189: contig of 2936 bp in length

78289: gap of unknown length

81876: contig of 3587 bp in length

81976: gap of unknown length

85415: contig of 3439 bp in length

85515: gap of unknown length

89541: contig of 4026 bp in length

89641: gap of unknown length

91906: contig of 2265 bp in length

92006: gap of unknown length

96061: contig of 4055 bp in length

96161: gap of unknown length

99759: contig of 3598 bp in length

99859: gap of unknown length

103561: contig of 3702 bp in length

103661: gap of unknown length

106540: contig of 2879 bp in length

106640: gap of unknown length

110562: contig of 3922 bp in length

110662: gap of unknown length

114866: contig of 4204 bp in length

114966: gap of unknown length

118735: contig of 3769 bp in length

118835: gap of unknown length

123824: contig of 4989 bp in length

123924: gap of unknown length

130816: contig of 6892 bp in length

130916: gap of unknown length

138536: contig of 7620 bp in length

Query Match 7.08; Score 21; DB 2; Length 180560;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 TTGGTGGAAAGGATGAACGA 78
 |||||
 Db 8999 TTGGTGGAAAGGATGAACGA 9019

RESULT 15
 AC094439
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-4E23, *** SEQUENCING IN PROGRESS ***,
 64 unordered pieces.
 AC094439
 VERSION AC094439.3 GI:21716480
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (bases 1 to 180989)
 Murny,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,
 Albrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbraia,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Burch,C., Burch,P., Burkett,C., Burrell,K.D., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
 Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
 Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R., Thomas,S.,
 Usman,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 180989)
 Worley,K.C.
 Direct Submission
 Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 180989)

AUTHORS
 TITLE
 JOURNAL
 COMMENT

Worley, K.C.
 Direct Submission
 Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 9, 2002 this sequence version replaced gi:17941167.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GAPK
 Center clone name: CH230-4E23
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 135185 bases at least Q40
 Consensus quality: 140034 bases at least Q30
 Consensus quality: 143806 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 64 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1133: contig of 1133 bp in length
 1233: gap of unknown length
 1234 2698: contig of 1465 bp in length
 2799 4428: gap of unknown length
 4429 4528: contig of 1630 bp in length
 4529 5919: gap of unknown length
 5920 6019: contig of 1391 bp in length
 6020 7277: gap of unknown length
 7277 8620: contig of 1257 bp in length
 8620 10136: contig of 1243 bp in length
 10136 10236: gap of unknown length
 10236 11871: contig of 1416 bp in length
 11871 13421: contig of 1635 bp in length
 13421 13521: gap of unknown length
 13521 14868: contig of 1451 bp in length
 14868 14968: gap of unknown length
 14968 16393: contig of 1347 bp in length
 16393 16493: gap of unknown length
 16493 17923: contig of 1425 bp in length
 17923 18023: gap of unknown length
 18023 19329: contig of 1430 bp in length
 19329 19429: gap of unknown length
 19429 22068: contig of 1306 bp in length
 22068 22168: gap of unknown length
 22168 24007: contig of 2639 bp in length
 24007 24107: gap of unknown length
 24107 25927: contig of 1839 bp in length
 25927 26027: gap of unknown length
 26027 27886: contig of 1820 bp in length
 27886 30208: contig of 1859 bp in length
 30208 30308: gap of unknown length
 30308 31676: contig of 2222 bp in length
 31676 33246: contig of 1367 bp in length
 33246 33347: gap of unknown length
 33347 35540: contig of 1471 bp in length
 35540 35540: gap of unknown length
 35540 35540: contig of 2194 bp in length

* 35541 35640: gap of unknown length
* 35641 37901: contig of 2261 bp in length
* 37902 38001: gap of unknown length
* 38002 39304: contig of 1303 bp in length
* 39305 39404: gap of unknown length
* 39405 41114: contig of 1710 bp in length
* 41115 41214: gap of unknown length
* 41215 42964: contig of 1750 bp in length
* 42965 43064: gap of unknown length
* 43065 45132: contig of 2068 bp in length
* 45133 45232: gap of unknown length
* 45233 46663: contig of 1431 bp in length
* 46664 46763: gap of unknown length
* 46764 49235: contig of 2472 bp in length
* 49236 49335: gap of unknown length
* 49336 51108: contig of 1773 bp in length
* 51109 51208: gap of unknown length
* 51209 54225: contig of 3017 bp in length
* 54226 54325: gap of unknown length
* 54326 56764: contig of 2439 bp in length
* 56765 56864: gap of unknown length
* 56865 58656: contig of 1792 bp in length
* 58657 58756: gap of unknown length
* 58757 61648: contig of 2892 bp in length
* 61649 61748: gap of unknown length
* 61749 64205: contig of 2457 bp in length
* 64206 64305: gap of unknown length
* 64306 67082: contig of 2777 bp in length
* 67083 67182: gap of unknown length
* 67183 70076: contig of 2894 bp in length
* 70077 70176: gap of unknown length
* 70177 74226: contig of 4050 bp in length
* 74227 74326: gap of unknown length
* 74327 77319: contig of 2993 bp in length
* 77320 77419: gap of unknown length
* 77420 80293: contig of 2874 bp in length
* 80294 80393: gap of unknown length
* 80394 82745: contig of 2352 bp in length
* 82746 82845: gap of unknown length
* 82846 84868: contig of 2023 bp in length
* 84869 84968: gap of unknown length
* 84969 89256: contig of 4288 bp in length
* 89257 89356: gap of unknown length
* 89357 92385: contig of 3029 bp in length
* 92386 92485: gap of unknown length
* 92486 95198: contig of 2713 bp in length
* 95199 95298: gap of unknown length
* 95299 98978: contig of 3680 bp in length
* 98979 99079: gap of unknown length
* 99079 102461: contig of 3383 bp in length
* 102462 102561: gap of unknown length
* 102562 105697: contig of 3136 bp in length
* 105698 105797: gap of unknown length
* 105798 108706: contig of 2909 bp in length
* 108707 108806: gap of unknown length
* 108807 111988: contig of 3182 bp in length
* 111989 112088: gap of unknown length
* 112089 114714: contig of 2626 bp in length
* 114715 114814: gap of unknown length
* 114815 118638: contig of 3824 bp in length
* 118639 118738: gap of unknown length
* 118739 122752: contig of 4014 bp in length
* 122753 122852: gap of unknown length
* 122853 126741: contig of 3889 bp in length
* 126742 126841: gap of unknown length
* 126842 130841: contig of 4000 bp in length

Query Match 7.0%; Score 21; DB 2; Length 180989;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 58 TTGGTGGAAAGGAATGAAGCA 78
|||||

Db 107913 TTGGTGGAAAGGAATGAAGCA 107933

Search completed: January 21, 2003, 13:47:25
Job time : 1511.45 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 10:00:41 ; Search time 1163.66 Seconds
(without alignments)
6877.663 Million cell updates/sec

Title: US-10-060-830-6

Perfect score: 275

Sequence: 1 caacttcagttgtcagccc.....tttgaagatgatgtgtttt 275

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_on.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	275	100.0	2939	9	AF387547	AF387547 Homo sapi
2	275	100.0	3858	9	BC029658	BC029658 Homo sapi
3	275	100.0	6093	9	AB073146	AB073146 Homo sapi
4	275	100.0	145515	2	AC025661	AC025661 Homo sapi
5	275	100.0	192247	2	AC013497	AC013497 Homo sapi
6	275	100.0	193623	9	AC106728	AC106728 Homo sapi
7	51	18.5	51	6	AX157055	AX157055 Sequence
8	25	9.1	51	6	AX157056	AX157056 Sequence
9	23	8.4	2339	10	AF387549	AF387549 Rattus no
10	23	8.4	2642	10	AF387548	AF387548 Mus muscu
11	23	8.4	203102	2	AC126975	AC126975 Rattus no
12	21	7.6	149582	2	AC112188	AC112188 Homo sapi
13	21	7.6	155942	9	AC096555	AC096555 Homo sapi
14	21	7.6	164290	2	AC096454	AC096454 Rattus no
15	21	7.6	180745	9	AL359183	AL359183 Human DNA
16	21	7.6	185319	2	AC110364	AC110364 Rattus no
17	20	7.3	28000	1	BSPBSXSE	Z70177 B.subtilis
18	20	7.3	39535	9	AC000072	AC000072 Homo sapi
19	20	7.3	47500	9	HUMDGCRT	L77569 Homo sapien
20	20	7.3	127339	10	AC000096	AC000096 Mus muscu
21	20	7.3	175878	2	AC121975	AC121975 Mus muscu
22	20	7.3	187274	2	AL391096	AL391096 Homo sapi
23	20	7.3	187470	2	AC087802	AC087802 Mus muscu
24	20	7.3	206880	2	AC073604	AC073604 Mus muscu
25	20	7.3	216750	1	BSUB0007	Z99110 Bacillus su
26	20	7.3	228345	10	AC079044	AC079044 Mus muscu
27	19	6.9	1059	10	BC025911	BC025911 Mus muscu
28	19	6.9	1654	3	AY075212	AY075212 Drosophill
29	19	6.9	3180	9	AK093825	AK093825 Homo sapi
30	19	6.9	60144	2	AC019798	AC019798 Drosophill
31	19	6.9	66254	9	AL137791	AL137791 Human DNA
32	19	6.9	70282	9	AP001505	AP001505 Homo sapi
33	19	6.9	85651	10	AL672144	AL672144 Mouse DNA
34	19	6.9	91685	2	AC005121	AC005121 Drosophill
35	19	6.9	107689	9	AL354681	AL354681 Human DNA
36	19	6.9	112584	8	ATT27115	AL358732 Arabidops
37	19	6.9	113681	9	HS1141E15	AL034422 Human DNA
38	19	6.9	114281	2	AC123107	AC123107 Rattus no
39	19	6.9	124888	9	AF001824	AF001824 Homo sapi
40	19	6.9	134970	9	AC002504	AC002504 Human Xp2
41	19	6.9	138033	2	AP002339	AP002339 Homo sapi
42	19	6.9	138244	9	AC004884	AC004884 Homo sapi
43	19	6.9	144302	9	AC091613	AC091613 Homo sapi
44	19	6.9	152037	9	HS167A19	AL031427 Human DNA
45	19	6.9	153356	2	AC004995	AC004995 Homo sapi

ALIGNMENTS

RESULT 1
AF387547
LOCUS AF387547
DEFINITION Homo sapiens endothelial and smooth muscle cell-derived
neuropilin-like protein (ESDN) mRNA, complete cds.
ACCESSION AF387547
VERSION AF387547.1 GI:16902434
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2939)
AUTHORS Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N.,
Matsumori,A., Sasayama,S., Honjo,T. and Tashiro,K.,

AF387547 2939 bp mRNA linear PRI 11-NOV-2001
Homo sapiens endothelial and smooth muscle cell-derived
neuropilin-like protein (ESDN) mRNA, complete cds.

TITLE ESDN, a novel neuropilin-like membrane protein cloned from vascular cells with the longest secretory signal sequence among eukaryotes, is up-regulated after vascular injury

J. Biol. Chem. 276 (36), 34105-34114 (2001)

21424027

11447234

2 (bases 1 to 2939)

Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N., Sasayama, S., Matsumori, A., Honjo, T. and Tashiro, K.

Direct Submission

Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

Location/Qualifiers

1. .2939

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="3"

1. .2939

/gene="ESDN"

12. .2339

/gene="ESDN"

/codon_start=1

/product="endothelial and smooth muscle cell-derived neuropilin-like protein"

/protein_id="AAL30178.1"

/db_xref="GI:16902435"

/translation="MASHAVVARRCPQCPVRAAARAAWAAALPLSLPPSCNSSS FEMFLFLLLVLLLEDDAGAQGGCGHTVLGPESGTLTSINPTYPNVCVME IRVAMGERVIRKFGDFIEDSDSCGHVIRGIVGSRTEIGVSGGLGOMNHSIE KNEITLFLMSGIHVSGFLASVVDKDLITCLDTASNFLEPFSESKYCPACGLP FAESGITPHGRDSSPLCMAGVHAGVYNTLGGQISVLSKIPYVESLANNVTSP VGLHSLTFETKTCGCTGLMGESGVADPOITASSVLEWTDHGOENSWKPKARLK KPGPPWAAFADEYQWLQIDLNKEKKITGITTCGTMVHNYVVSARILYSDGQKW TVYREPGEQDKIFGGKDYHQDVNNELPLIARFIRVNPQTQOQKIAMKMLLGCQ FIPGRPKPLTPPPRNSDLKNTAPPKAKGAPKFTQPLQPSRNEFPATGETQ TASPDRINTTVPNTKDVAALVPLVMVLTLLILLCVHWRNRKKTKETD LPYWDRAWGMKQKQFLPAKADHEETPVRYSSSEVNHLSPREVTVLQADSAEYAP LVGIVTGVLHGRSTFKPEEGEAGYADLPYNSPGVEYHAYAEPLTIGPEYATPII MDMSGHPTTSVGQSTSTFKATGNOPPLVGTNTYLLSRDSCSSAQAAQVDTPKAGP GLPAPDELIVYQVPOSTQEVSGAGRGCECDVPEIL"

BASE COUNT 825 a 675 c 657 g 782 t

ORIGIN

Query Match 100.0%; Score 275; DB 9; Length 2939;

Best Local Similarity 100.0%; Pred. No. 2.7e-153;

Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAACCTTCAGTTGGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC 60

Db 2080 CAACCTTCAGTTGGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC 2139

QY 61 CACTAGTGGGAACCTTACATACATCTCTCCAGGACTGACAGCTGCTCCTCAGCCGAGG 120

Db 2140 CACTAGTGGGAACCTTACATACATCTCTCCAGGACTGACAGCTGCTCCTCAGCCGAGG 2199

QY 121 CCAGTATGATACCCGGAAGCTGGGAAGCCAGGCTACTCTCCGCCAGACGAATGGTGT 180

Db 2200 CCAGTATGATACCCGGAAGCTGGGAAGCCAGGCTACTCTCCGCCAGACGAATGGTGT 2259

QY 181 ACCAGTGGCCAGACGACACAGAAGTATCAGGAGCAGGAAGGATGGGAATGTGATG 240

Db 2260 ACCAGTGGCCAGACGACACAGAAGTATCAGGAGCAGGAAGGATGGGAATGTGATG 2319

QY 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275

Db 2320 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 2354

RESULT 2

BC029658

LOCUS

DEFINITION Homo sapiens, similar to endothelial and smooth muscle cell-derived

neuropilin-like protein, clone MGC:30023 IMAGE:4431275, mRNA, complete cds.

BC029658

BC029658.1 GI:20988614

MGC.

ORGANISM Homo sapiens.

Homosapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3858)

Strausberg, R.

Direct Submission

Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-r@mail.nih.gov

Tissue Procurement: DCTD/BTP

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowals, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 42 Row: c Column: 3.

Location/Qualifiers

1. .3858

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clones="MGC:30023 IMAGE:4431275"

/tissue_type="Prostate, adenocarcinoma."

/clone_lib="NIH_MGC_91"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

21. .2252

/codon_start=1

/product="Similar to endothelial and smooth muscle cell-derived neuropilin-like protein"

/protein_id="AAH29658.1"

/translation="MPFLFLLLVLLLEDDAGAQGGCGHTVLGPESGTLTSINYP QTPNSTVCWEIRKMGERVIRKFGDFIEDSDSCGHVIRGIVGSRTEIGKYC GLGOMNHSIESKNEITLFLMSGIHVSGFLASVVDKDLITCLDTASNFLEPE FSKYCPAGCLLPFAEISGITPHGRDSSPLCMAGVHAGVYNTLGGQISVLSKIPY YESSLANNVTSPVGLHSLTFETKTCGCTGLMGESGVADPOITASSVLEWTDHGO ENSWKPKARLKFGPPWAAFADEYQWLQIDLNKEKKITGITTCGTMVHNYVVSAR ILYSDGQKWTVYREPGEQDKIFGGKDYHQDVNNELPLIARFIRVNPQTQOQK IAKMMLLGCQFIPGRPKPLTPPPRNSDLKNTAPPKAKGAPKFTQPLQPSR NEFPATGETQ TASPDRINTTVPNTKDVAALVPLVMVLTLLILLCVHWRNRKKTKETD SSEPYNHLSPREVTVLQADSAEYAPLVGIVTGVLHGRSTFKPEEGEAGYADLPY NSPGVEYHAYAEPLTIGPEYATPII MDMSGHPTTSVGQSTSTFKATGNOPPLVGTNTYLLSRDSCSSAQAAQVDTPKAGP GLPAPDELIVYQVPOSTQEVSGAGRGCECDVPEIL"

BASE COUNT 1137 a 807 c 785 g 1129 t

ORIGIN

Query Match 100.0%; Score 275; DB 9; Length 3858;

Best Local Similarity 100.0%; Pred. No. 2.7e-153;

Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAACCTTCAGTTGGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC 60

Db 1993 CAACCTTCAGTTGGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC 2052

[illegible]

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Melclm, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 146515)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bida, F.,
 Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domino, M., Doyle, M., Ferrel, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Melclm, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 12, 2000 this sequence version replaced gi:7230271.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE JOURNAL COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7942
 Center clone name: 322_A_20

* NOTE: This record contains 164 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

* 1. 836: contig of 836 bp in length
 * 837 936: gap of 100 bp
 * 937 1748: contig of 812 bp in length
 * 1749 1848: gap of 100 bp
 * 1849 2680: contig of 832 bp in length
 * 2681 2780: gap of 100 bp
 * 2781 3534: contig of 754 bp in length

* 3535 3634: gap of 100 bp
 * 3635 4383: contig of 749 bp in length
 * 4384 4483: gap of 100 bp
 * 4484 5261: contig of 778 bp in length
 * 5262 5361: gap of 100 bp
 * 5362 6136: contig of 775 bp in length
 * 6137 6236: gap of 100 bp
 * 6237 7045: contig of 809 bp in length
 * 7046 7145: gap of 100 bp
 * 7146 7953: contig of 808 bp in length
 * 7954 8053: gap of 100 bp
 * 8054 8857: contig of 804 bp in length
 * 8858 8957: gap of 100 bp
 * 8958 9784: contig of 827 bp in length
 * 9785 9884: gap of 100 bp
 * 9885 10718: contig of 834 bp in length
 * 10719 10818: gap of 100 bp
 * 10819 11640: contig of 822 bp in length
 * 11641 11740: gap of 100 bp
 * 11741 12534: contig of 794 bp in length
 * 12535 12634: gap of 100 bp
 * 12635 13407: contig of 773 bp in length
 * 13408 13507: gap of 100 bp
 * 13508 14276: contig of 769 bp in length
 * 14277 14376: gap of 100 bp
 * 14377 15185: contig of 809 bp in length
 * 15186 15285: gap of 100 bp
 * 15285 16086: contig of 801 bp in length
 * 16087 16186: gap of 100 bp
 * 16187 16986: contig of 800 bp in length
 * 16987 17086: gap of 100 bp
 * 17087 17913: contig of 827 bp in length
 * 17914 18013: gap of 100 bp
 * 18014 18841: contig of 828 bp in length
 * 18842 18941: gap of 100 bp
 * 18942 19770: contig of 829 bp in length
 * 19771 19870: gap of 100 bp
 * 19871 20689: contig of 819 bp in length
 * 20690 20789: gap of 100 bp
 * 20790 21575: contig of 786 bp in length
 * 21576 21675: gap of 100 bp
 * 21676 22423: contig of 748 bp in length
 * 22424 22523: gap of 100 bp
 * 22524 23322: contig of 799 bp in length
 * 23323 23422: gap of 100 bp
 * 23423 24239: contig of 817 bp in length
 * 24240 24339: gap of 100 bp
 * 24340 25199: contig of 860 bp in length
 * 25200 25299: gap of 100 bp
 * 25300 26131: contig of 832 bp in length
 * 26132 26231: gap of 100 bp
 * 26232 27103: contig of 872 bp in length
 * 27104 27203: gap of 100 bp
 * 27204 28060: contig of 857 bp in length
 * 28061 28160: gap of 100 bp
 * 28161 29013: contig of 853 bp in length
 * 29014 29113: gap of 100 bp
 * 29114 29944: contig of 831 bp in length
 * 29945 30044: gap of 100 bp
 * 30045 30746: contig of 702 bp in length
 * 30747 30846: gap of 100 bp
 * 30847 31652: contig of 806 bp in length
 * 31653 31752: gap of 100 bp
 * 31753 32559: contig of 807 bp in length
 * 32560 32659: gap of 100 bp
 * 32660 33470: contig of 811 bp in length
 * 33471 33570: gap of 100 bp
 * 33571 34386: contig of 816 bp in length
 * 34387 34486: gap of 100 bp
 * 34487 35290: contig of 804 bp in length
 * 35291 35390: gap of 100 bp
 * 35391 36199: contig of 809 bp in length
 * 36200 36299: gap of 100 bp

JOURNAL
REFERENCE
AUTHORS

Unpublished 2 (bases 1 to 192247)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Doyle, M.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6692347.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Center project name: L1852
Center clone name: 319_J-24
----- Summary Statistics -----

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; version 0.960731
Assembly program: Phrap; version 0.960731
Consensus quality: 182870 bases at least Q40
Consensus quality: 187032 bases at least Q30
Consensus quality: 188925 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 191247; sum-of-contigs
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 11 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 472: contig of 472 bp in length
* 473 572: gap of 100 bp
* 573 3945: contig of 3373 bp in length
* 3946 4045: gap of 100 bp
* 4046 11056: contig of 7011 bp in length
* 11057 11156: gap of 100 bp
* 11157 17728: contig of 6572 bp in length
* 17729 17828: gap of 100 bp
* 17829 30556: contig of 12728 bp in length
* 30557 30656: gap of 100 bp
* 30657 49294: contig of 18638 bp in length
* 49295 49394: gap of 100 bp
* 49395 64509: contig of 15115 bp in length
* 64510 64609: gap of 100 bp
* 64610 83042: contig of 18433 bp in length
* 83043 83142: gap of 100 bp
* 83143 112373: contig of 29231 bp in length
* 112374 112473: gap of 100 bp
* 112474 145663: contig of 33190 bp in length
* 145664 145763: gap of 100 bp
* 145764 192247: contig of 46484 bp in length.

Location/Qualifiers

FEATURES

36300 37156: contig of 857 bp in length
* 37157 37256: gap of 100 bp
* 37257 38097: contig of 841 bp in length
* 38098 38197: gap of 100 bp
* 38198 39015: contig of 818 bp in length
* 39016 39115: gap of 100 bp
* 39116 39895: contig of 780 bp in length
* 39896 39995: gap of 100 bp
* 39996 40786: contig of 791 bp in length
* 40787 40886: gap of 100 bp
* 40887 41686: contig of 800 bp in length
* 41687 41786: gap of 100 bp
* 41787 42549: contig of 763 bp in length
* 42550 42649: gap of 100 bp
* 42650 43459: contig of 810 bp in length
* 43460 43559: gap of 100 bp
* 43560 44293: contig of 734 bp in length
* 44294 44393: gap of 100 bp
* 44394 45225: contig of 832 bp in length
* 45226 45325: gap of 100 bp
* 45326 46138: contig of 813 bp in length
* 46139 46238: gap of 100 bp
* 46239 46987: contig of 749 bp in length
* 46988 47087: gap of 100 bp
* 47088 47853: contig of 766 bp in length
* 47854 47953: gap of 100 bp
* 47954 48758: contig of 805 bp in length
* 48759 48858: gap of 100 bp
* 48859 49664: contig of 806 bp in length
* 49665 49764: gap of 100 bp
* 49765 50569: contig of 805 bp in length
* 50570 50669: gap of 100 bp

Query Match 100.0%; Score 275; DB 2; Length 146515;
Best Local Similarity 100.0%; Pred. No. 3.1e-153;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAACCTCAGTTGGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC 60
Db 48105 CAACCTCAGTTGGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC 48164

Qy 61 CACTAGTGGGAACCTTACAACTCTCTCCAGGACTGACAGCTGCTCCTCAGCCAGG 120
Db 48165 CACTAGTGGGAACCTTACAACTCTCTCCAGGACTGACAGCTGCTCCTCAGCCAGG 48224

Qy 121 CCCAGTATGATACCCGAACTGGGAGCCAGGCTCTACCTGCCCCACACGAATTTGGTGT 180
Db 48225 CCCAGTATGATACCCGAACTGGGAGCCAGGCTCTACCTGCCCCACACGAATTTGGTGT 48284

Qy 181 ACCAGTGCCACAGACGACACAAAGATATCAGGAGGAGGATGGGGAATGTGATG 240
Db 48285 ACCAGTGCCACAGACGACACAAAGATATCAGGAGGAGGATGGGGAATGTGATG 48344

Qy 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275
Db 48345 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 48379

RESULT 5
AC013497
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-319J24 map 3, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC013497
AC013497.4 GI:7329299
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 192247)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 3, clone RP11-319J24

```
source
1. .192247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-319J24"
/clone_lib="RPC1-11 Human Male BAC"
1. .472
/clone="assembly_fragment"
/clone_end:SP6
vector_side:right
573. .3945
/clone="assembly_fragment"
4046. .11056
/clone="assembly_fragment"
11157. .17728
/clone="assembly_fragment"
17829. .30556
/clone="assembly_fragment"
30657. .49294
/clone="assembly_fragment"
49395. .64509
/clone="assembly_fragment"
64610. .83042
/clone="assembly_fragment"
clone_end:T7
vector_side:left
83143. .112373
/clone="assembly_fragment"
112474. .145863
/clone="assembly_fragment"
145764. .192247
/clone="assembly_fragment"

BASE COUNT 58524 a 38157 c 36640 g 57721 t 1005 others
ORIGIN

Query Match 100.0%; Score 275; DB 2; Length 192247;
Best Local Similarity 100.0%; Pred. No. 3.1e-153;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAACCTCAGTGGTTCAGCCCTCCACATCCACTTTCAGGCTACGGGGACCAACTCCCC 60
Db 165188 CAACCTCAGTGGTTCAGCCCTCCACATCCACTTTCAGGCTACGGGGACCAACTCCCC 165247

QY 61 CACTAGTGGGAACCTTACATACACTTCTCCAGCACTGACAGTGTCTCTCCAGCCAGG 120
Db 165248 CACTAGTGGGAACCTTACATACACTTCTCTCCAGCACTGACAGTGTCTCTCCAGCCAGG 165307

QY 121 CCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCGACAGCAATTGTGT 180
Db 165308 CCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCGACAGCAATTGTGT 165367

QY 181 ACCAGTCCACAGCAGCACAGAAGTATCAGGAGGAGGATGGGGAATTGTGATG 240
Db 165368 ACCAGTCCACAGCAGCACAGAAGTATCAGGAGGAGGATGGGGAATTGTGATG 165427

QY 241 TTTTAAAGAATCTTTTGAAGATGATGCTGCTTT 275
Db 165428 TTTTAAAGAATCTTTTGAAGATGATGCTGCTTT 165462

RESULT 6
AC106728
LOCUS AC106728 Homo sapiens 3 BAC RP11-319J24 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC106728
VERSION AC106728.6 GI:21206081
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE
AUTHORS

```
1 (bases 1 to 193623)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryan,N.P.,
Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisege,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Morris,S., Moser,M., Neal,D.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenwo,S., Ogih,M., Okwuonu,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shooshkari,N., Sisson,I., Sodergren,E.,
Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Villalon,D., Vinson,R., Wang,O., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
Wu,Y.F., Zhou,J., Zorrilla,S., Zylor,S.L., Weinstein,G. and
Gibbs,R.
```

TITLE
JOURNAL

Unpublished

2 (bases 1 to 193623)

Worley,K.C.

TITLE
JOURNAL

Direct Submission

Submitted (12-JAN-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 193623)

Worley,K.C.

TITLE
JOURNAL

Direct Submission

Submitted (09-MAY-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 193623)

Worley,K.C.

TITLE
JOURNAL

Direct Submission

Submitted (25-MAY-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On May 25, 2002 this sequence version replaced gi:20429236.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the

entire insert of this clone. Overlapping regions of clones are only

sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the

Features listing.

COMMENT

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Res. of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES	source	Location/Qualifiers
		1. 193623
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="3"
		/clone="RP11-319J24"
repeat_region		818..1088
repeat_region		/rpt_family="L2"
repeat_region		complement(3791..3934)
repeat_region		/rpt_family="MIR"
repeat_region		complement(4203..4581)
repeat_region		/rpt_family="L2"
repeat_region		4767..4790
repeat_region		/rpt_family="(TC)n"
repeat_region		4791..4851
repeat_region		/rpt_family="(TC)n"
repeat_region		4854..4892
repeat_region		/rpt_family="(CA)n"
repeat_region		4907..5066
repeat_region		/rpt_family="L2"
repeat_region		5333..5386
repeat_region		/rpt_family="GA-rich"
repeat_region		complement(5512..5945)
repeat_region		/rpt_family="MLTIC"
repeat_region		complement(6187..6624)
repeat_region		/rpt_family="LTR37B"
repeat_region		complement(6858..7219)
repeat_region		/rpt_family="L2"
STS		6880..7000
repeat_region		/standard_name="73424"
repeat_region		complement(7409..7632)
repeat_region		/rpt_family="MIR"
repeat_region		7862..8369
repeat_region		/rpt_family="L1MC1"
repeat_region		8417..8524
repeat_region		/rpt_family="L1M4"
repeat_region		complement(10026..10215)
repeat_region		/rpt_family="MER20"
repeat_region		complement(10485..10549)
repeat_region		/rpt_family="MER112"
repeat_region		12814..13590
repeat_region		/rpt_family="L1PA7"
repeat_region		13917..14135
repeat_region		/rpt_family="L1MA5A"
repeat_region		complement(14753..15070)

repeat_region	/rpt_family="AluSq"	15152..15366
repeat_region	/rpt_family="MIR"	15395..15562
repeat_region	/rpt_family="MER5B"	16109..16388
repeat_region	/rpt_family="L1MC5"	16389..16642
repeat_region	/rpt_family="AluY"	16643..16758
repeat_region	/rpt_family="L1MC5"	16826..16898
repeat_region	/rpt_family="L1MC5"	complement(17591..18655)
repeat_region	/rpt_family="L1MC4"	complement(19191..19653)
repeat_region	/rpt_family="L1MD3"	complement(19662..19780)
repeat_region	/rpt_family="FLAM_A"	complement(19781..21049)
repeat_region	/rpt_family="L1MD3"	complement(21069..21455)
repeat_region	/rpt_family="MLTIB"	complement(21479..22318)
repeat_region	/rpt_family="L1MD3"	complement(22447..22681)
repeat_region	/rpt_family="L1MD3"	complement(23470..23853)
repeat_region	/rpt_family="MSTD"	23852..23925
repeat_region	/rpt_family="U3"	23936..25203
repeat_region	/rpt_family="L1PB3"	24844..24975
STS	/standard_name="6234"	25219..25388
repeat_region	/rpt_family="(TTTC)n"	complement(25428..25548)
repeat_region	/rpt_family="AluJb"	25555..25950
repeat_region	/rpt_family="L1PB3"	25951..26456
repeat_region	/rpt_family="MER75"	26457..27565

Query Match 100.0%; Score 275; DB 9; Length 193623;
 Best Local Similarity 100.0%; Pred. No. 3.1e-153; Indels 0; Gaps 0;
 Matches 275; Conservative 0; Mismatches 0;

Qy	1	CAACTTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC	60
Db	96401	CAACTTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC	96460
Qy	61	CACATAGTGGGAACCTTACAAATACACTTCTCCAGACTGACAGCTGCTCTCAGCCAGG	120
Db	96461	CACATAGTGGGAACCTTACAAATACACTTCTCCAGACTGACAGCTGCTCTCAGCCAGG	96520
Qy	121	CCCAGTATGATACCCCGAAAGCTGGGAGCCAGGCTACCTGCCAGACCAATGGTGT	180
Db	96521	CCCAGTATGATACCCCGAAAGCTGGGAGCCAGGCTACCTGCCAGACCAATGGTGT	96580
Qy	181	ACCAGTGCACAGACGACACAAAGATATCAGGACGAGGATGGGAATGTGATG	240
Db	96581	ACCAGTGCACAGACGACACAAAGATATCAGGACGAGGATGGGAATGTGATG	96640
Qy	241	TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT	275
Db	96641	TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT	96675

RESULT 7
 AX157055
 LOCUS

AX157055

51 bp DNA

linear PAT 22-JUN-2001

```

DEFINITION Sequence 383 from Patent WO0140521.
ACCESSION AX157055
VERSION AX157055.1 GI:14538386
KEYWORDS human.
SOURCE ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 51)
AUTHORS Shimkets,R.A. and Leach,M.
TITLE Nucleic acids containing single nucleotide polymorphisms and
methods of use thereof
JOURNAL Patent: WO 0140521-A 383 07-JUN-2001;
Curagen Corporation (US)
FEATURES
source
1..51
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
misc_feature
26
/feature="1 of 2 allelic variants (384 is other entry)"
Accession number cg44922119
BASE COUNT 13 a 17 c 14 g 7 t
ORIGIN
Query Match 18.5%; Score 51; DB 6; Length 51;
Best Local Similarity 100.0%; Pred. No. 6.4e-19;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 CAGCCAGCCAGCCAGTATGATACCCGAGAGCTGGGAGCGAGTCTACTG 162
|||||
Db 1 CAGCCAGCCAGCCAGTATGATACCCGAGAGCTGGGAGCGAGTCTACTG 51

RESULT 8
AX157056
LOCUS AX157056
DEFINITION Sequence 384 from Patent WO0140521.
ACCESSION AX157056
VERSION AX157056.1 GI:14538387
KEYWORDS human.
SOURCE ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 51)
AUTHORS Shimkets,R.A. and Leach,M.
TITLE Nucleic acids containing single nucleotide polymorphisms and
methods of use thereof
JOURNAL Patent: WO 0140521-A 384 07-JUN-2001;
Curagen Corporation (US)
FEATURES
source
1..51
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
misc_feature
26
/feature="2 of 2 allelic variants (383 is other entry)"
Accession number cg44922119
BASE COUNT 14 a 17 c 13 g 7 t
ORIGIN
Query Match 9.1%; Score 25; DB 6; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 CAGCCAGCCAGCCAGTATGATACCC 136
|||||
Db 1 CAGCCAGCCAGCCAGTATGATACCC 25

RESULT 9
AF387549
LOCUS AF387549
DEFINITION Rattus norvegicus endothelial and smooth muscle cell-derived
Sequence 2339 bp mRNA linear ROD 11-NOV-2001
ACCESSION AF387549
VERSION AF387549.1 GI:16902436
KEYWORDS Rattus norvegicus.
SOURCE ORGANISM Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2339)
AUTHORS Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N.,
Matsumori,A., Sasayama,S., Honjo,T. and Tashiro,K.
TITLE ESDN, a novel neuropilin-like membrane protein cloned from vascular
cells with the longest secretory signal sequence among eukaryotes,
is up-regulated after vascular injury
JOURNAL J. Biol. Chem. 276 (36), 34105-34114 (2001)
FEATURES
source
1..2339
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1..2339
/gene="Esdn"
12..2321
/gene="Esdn"
/codon_start=1
/product="endothelial and smooth muscle cell-derived
neuropilin-like protein"
/protein_id="AAL30180.1"
/db_xref="GI:16902439"
/translation="MASRAPLRARSPODGRAPAPATGRAPLPSAGWCPLPPGRNS
SSRPLRLLLLLPDAQAQDGGCHTVLGPESGRTLSINYPSTVCKREIRV
KTGRIRIKFGDFDIEDSDCHNLNLYKFNIGYVSRTEIGYCGLGQNMQISKSGS
EITVLFMSGIIHSGRGFLASVVDKQDLTCLDTNFKLPESKYCPAGCLLPFAE
ISGIIPIHGIRDSPLCMAGIHAGVSDVLGGQISVVISKGTPIYESSIANNVTSNVGY
LSTSLFTKTSKGCYTLGMESEVIADPOITASSVLEHTDNGQENSWKPEKARLKPG
PPWAFATDEHQWLQIDLNKKKIKGIVTGTSLIEHNYISATRVLISDDGQWTVI
REFGAQDKIFQGNRDKHDKVNNFLPPIIARFIRNVPVQQAAMVELLGGCOFTL
KGRPLKLTQPPPPNNNNKNTVHPKIGRAPKFTQALQPRSRNDLPLLPQTATPD
VKNTVTPTSVTKOVALAALVPLVLMALTTLLILVCAWHNRNRKKAEGTYDLPHPD
RAGWVGKVKOLLPAKSVSEHEETPVYNSSEVSHLSPREVTVLQADSAEYAQPLVGGI
VGTLHORSTFKPEEGKEASVADLPYNAPQVEYHVAEPLPTVGPVATPIVNDMSG
HSTASVGLPSTSTPTAGNQPPALVGTNTLLSTDSQSSQAQYDTPKGGKPAARPE
ELIYQVPSQEQEASGAGNDFKFAKFTL"
BASE COUNT 612 a 606 c 594 g 527 t
ORIGIN
Query Match 8.4%; Score 23; DB 10; Length 2339;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CTTCTCTCCAGGACTGACAGCTG 106
|||||
Db 2148 CTTCTCTCCAGGACTGACAGCTG 2170

RESULT 10
AF387548
LOCUS AF387548
DEFINITION Mus musculus strain ICR endothelial and smooth muscle cell-derived
neuropilin-like protein (Esdn) mRNA, complete cds.
ACCESSION AF387548
VERSION AF387548.1 GI:16902436
KEYWORDS

```

```

SOURCE      Mus musculus.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 2642)
AUTHORS     Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,
            Matsumori, A., Sasayama, S., Honjo, T. and Tashiro, K.
TITLE       ESDN, a novel neuropilin-like membrane protein cloned from vascular
            cells with the longest secretory signal sequence among eukaryotes,
            is up-regulated after vascular injury
JOURNAL     J Biol Chem. 276 (36), 34105-34114 (2001)
MEDLINE     21424027
PUBMED      11447234
REFERENCE   2 (bases 1 to 2642)
AUTHORS     Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,
            Sasayama, S., Matsumori, A., Honjo, T. and Tashiro, K.
TITLE       Direct Submision
JOURNAL     Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto
            University Graduate School of Medicine, Yoshida Konoe-cho,
            Sakyo-ku, Kyoto 606-8501, Japan
FEATURES    Location/Qualifiers
             1..2642
                /organism="Mus musculus"
                /strain="ICR"
                /db_xref="taxon:10090"
             1..2642
                /gene="Esdn"
                /note="similar to the Mus musculus strain 129/SvJ Esdn
                gene sequence in GenBank Accession Number AF442820"
             314..2623
                /gene="Esdn"
                /codon_start=1
                /product="endothelial and smooth muscle cell-derived
                neuropilin-like protein"
                /protein_id="AA130179.1"
                /db_xref="GI:16902437"
                /translation="MASRAFLRAARSPQGGPAAPATGRAALPSAGCCPLPPGRNS
                SSRRLILLILLDAGGQDGGCHTGLVPGSGTLTSINPHTYSPNVCWEIRV
                RTGRIRIKFGDFIEDSYCHLNKLFNGIGVSRTEIGKYGGLQMNGSIESKGS
                ETVFLNKGTHAAGRFLASVYSDIKEDLITCLDTSNLFEPSEKYPACGLLPFAE
                ISGIPHGRRDSSPLCMAGHAGVYVSNVLGGQISIVISKGTPYESSLANVSTVGY
                LSASLFTKTSQGLTGMESGVADPOITASSALEMDTHDMQNSWTAERAKRKPG
                PPAAAFADDEHQLIDLNKKEKITGTVTGTSTIEHSYVSAYRVLYSDGQWRVTI
                REPGVDQKIFQGNKHDKYRNFLPPIIARFRNVPVQWQKIAMVELLGGQFTL
                KGRPKLTTPPRGNLRNTTARPKLGRAPKFTQVLPQSRNELPQPAETTTPTD
                IKNTVTPSYTKDVALAALVPLVLMALTLLILVCAWHNRNKKKTEGAVDLPHWD
                RAGWKGKQLLPKAKSDHEETPVRYSTSEVSHLSAREVTIVLQADSREYQPLVGGI
                VGTLLHQSTFKPEGKEAGYADLDPPYNPMQEVHYAYAEPLPVTGPEYATPIVMDMSG
                HPTASVGLPSTSTFKTAGTOPHALVGTYNITLLSRDTSQSGAQYDTPKGGKSAATPE
                ELVYQVQSTQELSGAGRDEKFAKEIL"
             431..778
                /gene="Esdn"
                /note="corresponds to nt 453 to 800, referred to in J.
                Biol. Chem. 276 (36), 34105-34114 (2001)"
BASE COUNT 639 a 703 c 744 g 556 t
ORIGIN
misc_feature
Query Match      8.4%; Score 23; DB 10; Length 2642;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 CTTCTCTCCAGACTGACAGCTG 106
Db 2450 CTTCTCTCCAGACTGACAGCTG 2472
RESULT 11
AC126975/c
LOCUS      AC126975
DEFINITION Rattus norvegicus clone CH230-234N3, *** SEQUENCING IN PROGRESS
            *** 74 unorded pieces.
ACCESSION AC126975
VERSION   AC126975.1 GI:21731383

```

```

KEYWORDS    HTG; HTGS_PHASE1.
            Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 203102)
AUTHORS     Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
            Alsbrooks, S.I., Amarantunga, H.C., Are, J.R., Ayale, M., Banks, T.,
            Barbara, J., Benton, J., Blmage, K., Blankenburg, K., Bonnin, D.,
            Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
            Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
            Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
            Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
            Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
            Davilla, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
            Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
            Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
            Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
            Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
            Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
            Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
            Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
            Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
            Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
            Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
            Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
            Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
            Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
            Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
            Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
            Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
            Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
            Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
            Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
            Oragune, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
            Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
            Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savery, G.,
            Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
            Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
            Sutton, A., Syatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
            Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
            Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
            Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
            Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,
            Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
            Weinstein, G. and Gibbs, R.
            Direct Submission
            Unpublished
            2 (bases 1 to 203102)
            Worley, K.C.
            Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: GJZO
            Center clone name: CH230-234N3
            ----- Summary Statistics
            Sequencing vector: Plasmid;
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 145293 bases at least Q40
            Consensus quality: 154399 bases at least Q30
            Consensus quality: 161802 bases at least Q20
            -----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 74 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1310: contig of 1310 bp in length
1311 1410: gap of unknown length
1411 2497: contig of 1087 bp in length
2498 2597: gap of unknown length
2598 3718: contig of 1121 bp in length
3719 3818: gap of unknown length
3819 4974: contig of 1156 bp in length
4975 5074: gap of unknown length
5076 6118: contig of 1044 bp in length
6119 6218: gap of unknown length
6219 7276: contig of 1058 bp in length
7277 7376: gap of unknown length
7377 8394: contig of 1018 bp in length
8395 8494: gap of unknown length
8495 8999: contig of 1405 bp in length
8999 9999: gap of unknown length
10000 11010: contig of 1011 bp in length
11011 11110: gap of unknown length
11111 12429: contig of 1319 bp in length
12430 12529: gap of unknown length
12530 14233: contig of 1704 bp in length
14234 14333: gap of unknown length
14334 15599: contig of 1266 bp in length
15600 15699: gap of unknown length
15700 16726: contig of 1027 bp in length
16727 16826: gap of unknown length
16827 18918: contig of 2092 bp in length
18919 19018: gap of unknown length
19019 20837: contig of 1819 bp in length
20838 20937: gap of unknown length
20939 23036: contig of 2099 bp in length
23037 23136: gap of unknown length
23137 24890: contig of 1754 bp in length
24891 24990: gap of unknown length
24991 26683: contig of 1693 bp in length
26684 26783: gap of unknown length
26784 28087: contig of 1304 bp in length
28088 28187: gap of unknown length
28188 28847: contig of 1660 bp in length
28848 29947: gap of unknown length
29948 31165: contig of 1218 bp in length
31166 31265: gap of unknown length
31266 32327: contig of 1062 bp in length
32328 32427: gap of unknown length
32428 33554: contig of 1127 bp in length
33555 34694: gap of unknown length
34695 34794: gap of unknown length
34795 36507: contig of 1713 bp in length
36508 36607: gap of unknown length
36608 38322: contig of 1715 bp in length
38323 38422: gap of unknown length
38423 40396: contig of 1974 bp in length
40397 40496: gap of unknown length
40497 42695: contig of 2199 bp in length
42696 42795: gap of unknown length
42796 43853: contig of 1058 bp in length
43854 43954: gap of unknown length
43954 45008: contig of 1055 bp in length
45009 45108: gap of unknown length
45109 47026: contig of 1918 bp in length
47027 47126: gap of unknown length
47127 49195: contig of 2069 bp in length
49196 49295: gap of unknown length
49296 51748: contig of 2453 bp in length

```

```

51749 51848: gap of unknown length
51849 54202: contig of 2354 bp in length
54203 54302: gap of unknown length
54303 5606: contig of 2304 bp in length
5607 56706: gap of unknown length
56707 58300: contig of 1594 bp in length
58301 58400: gap of unknown length
58401 59288: contig of 1528 bp in length
59289 60028: gap of unknown length
60029 62227: contig of 2199 bp in length
62228 62327: gap of unknown length
62328 64445: contig of 2118 bp in length
64446 64545: gap of unknown length
64546 66642: contig of 2097 bp in length
66643 66742: gap of unknown length
66743 68545: contig of 1803 bp in length
68546 68645: gap of unknown length
68646 70370: contig of 1725 bp in length
70371 70470: gap of unknown length
70471 72796: contig of 2326 bp in length
72797 72896: gap of unknown length
72897 75152: contig of 2256 bp in length
75153 75252: gap of unknown length
75253 77649: contig of 2397 bp in length
77650 77749: gap of unknown length
77750 80695: contig of 2946 bp in length
80696 80795: gap of unknown length
80796 83797: contig of 3002 bp in length
83798 83897: gap of unknown length
83898 86763: contig of 2866 bp in length
86764 86863: gap of unknown length
86864 89892: contig of 3029 bp in length
89893 89992: gap of unknown length
89993 92105: contig of 2113 bp in length
92106 92205: gap of unknown length
92206 94330: contig of 2125 bp in length
94331 94430: gap of unknown length
94431 96835: contig of 2405 bp in length
96836 96935: gap of unknown length
96936 98553: contig of 2618 bp in length
98554 99653: gap of unknown length
99654 102601: contig of 2948 bp in length
102602 102701: gap of unknown length
102702 105356: contig of 2655 bp in length
105357 105456: gap of unknown length
105457 109434: contig of 3378 bp in length
109435 109534: gap of unknown length
109535 112973: contig of 3439 bp in length
112974 113073: gap of unknown length

```

Query Match 8.4%; Score 23; DB 2; Length 203102;
 Best Local Similarity 100.0%; Pred. No. 0.055;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 CTTCTCTCCAGGACTGACAGCTG 106
 Db 179881 CTTCTCTCCAGGACTGACAGCTG 179859

RESULT 12

AC112188/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC112188 Homo sapiens chromosome 5 clone RP11-2L3, WORKING DRAFT SEQUENCE,
 11 unordered pieces.
 AC112188.1 GI:18767460
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
 Homo sapiens.
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 149582)
 DOE Joint Genome Institute.

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 149582)
DOE Joint Genome Institute.
Direct Submission
Submitted (20-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 407400
Center clone name: RPCI-11_2L3
-----
Summary Statistics
Consensus quality: 138239 bases at least Q40
Consensus quality: 142276 bases at least Q30
Consensus quality: 144090 bases at least Q20
Estimated insert size: 163000; agarose-fp estimation
Estimated insert size: 148582; sum-of-contigs estimation
Quality coverage: 3.98 in Q20 bases; agarose-fp estimation
Quality coverage: 4.37 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1344: contig of 1344 bp in length
* 1345 1444: gap of unknown length
* 1445 3792: contig of 2348 bp in length
* 3793 3892: gap of unknown length
* 3893 7155: contig of 3263 bp in length
* 7156 7255: gap of unknown length
* 7256 11627: contig of 4372 bp in length
* 11628 11727: gap of unknown length
* 11728 16849: contig of 5122 bp in length
* 16850 16949: gap of unknown length
* 16950 25439: contig of 8490 bp in length
* 25440 25539: gap of unknown length
* 25540 35440: contig of 9901 bp in length
* 35441 35540: gap of unknown length
* 35541 47044: contig of 11504 bp in length
* 47045 47144: gap of unknown length
* 47145 74054: contig of 26909 bp in length
* 74054 74153: gap of unknown length
* 74154 103413: contig of 29260 bp in length
* 103414 103513: gap of unknown length
* 103514 149582: contig of 46069 bp in length.
FEATURES
Location/Qualifiers
1..149582
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RPCI-11_2L3"
/clone_lib="RPCI human BAC library 11"
BASE COUNT 49371 a 26662 c 26162 g 46380 t 1007 others
ORIGIN

Query Match 7.6%; Score 21; DB 2; Length 149582;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 TGTGTTTAAAGAAATCCTTGG 259
|||||
Db 128313 TGTGTTTAAAGAAATCCTTGG 128293

RESULT 13

```

AC096555 155942 bp DNA linear PRI 01-MAR-2002
Homo sapiens BAC clone RP11-78N15 from 2, complete sequence.
AC096555 AC073455
AC096555.1 GI:15638740
HTG.
SOURCE
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155942)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 155942)
Shah, N., Kozlowicz, A. and Elliott, G.
The sequence of Homo sapiens BAC clone RP11-78N15
Unpublished (2001)
3 (bases 1 to 155942)
Waterston, R.H.
Direct Submission
Submitted (18-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 155942)
Waterston, R.
Direct Submission
Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 18, 2001 this sequence version replaced gi:8901261.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplense@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0078N15
Drafting Center: WUGSC

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-315C24; the clone sequenced

to the right is RP11-70G18, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-78N15.

Data from AC093639 was used to finish this clone, AC073455. Polymorphisms exist between RP11-78N15 and RP11-315C24.

The sequence of AC073455 has been incorporated into AC096555.

FEATURES

```

source
    location/Qualifiers
    1. .155942
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="2"
    /map="2"
    /clone="RP11-78N15"
    /clone_lib="RPCI-11"
    2. .45
    /rpt_family="MaLR"
    repeat_region
    2188..2215
    /rpt_family="AT_rich"
    misc_feature
    2230..2257
    /note="match to EST AW665990 (NID:g7458539) hj10g03.x1"
    repeat_region
    3253..3513
    /rpt_family="MaLR"
    repeat_region
    3520..3684
    /rpt_family="MaLR"
    repeat_region
    5795..5923
    /rpt_family="ERV1"
    repeat_region
    6957..6986
    /rpt_family="AT_rich"
    repeat_region
    7175..7485
    /rpt_family="Alu"
    repeat_region
    7456..7490
    /rpt_family="(A)n"
    repeat_region
    7825..8341
    /rpt_family="ERV1"
    repeat_region
    8952..8971
    /rpt_family="(T)n"
    repeat_region
    8954..9402
    /rpt_family="L1"
    repeat_region
    9519..9826
    /rpt_family="Alu"
    repeat_region
    9975..10028
    /rpt_family="T-rich"
    repeat_region
    10007..10315
    /rpt_family="Alu"
    repeat_region
    10422..10733
    /rpt_family="Alu"
    repeat_region
    10707..10737
    /rpt_family="(GAAA)n"
    repeat_region
    10863..11114
    /rpt_family="MIR"
    repeat_region
    11187..11223
    /rpt_family="(TTA)n"
    repeat_region
    11365..11393
    /rpt_family="(TTA)n"
    repeat_region
    11726..11910
    /rpt_family="MIR"
    repeat_region
    11911..12218
    /rpt_family="Alu"
    repeat_region
    12192..12218
    /rpt_family="(A)n"
    repeat_region
    12219..12267
    /rpt_family="MIR"
    repeat_region
    12358..12654
    /rpt_family="Alu"
    repeat_region
    13449..13605
    /rpt_family="MIR"
    repeat_region
    13606..14114
    /rpt_family="MER1_type"
    repeat_region
    14115..14185
    /rpt_family="MIR"
    repeat_region
    14242..14264
    /rpt_family="AT_rich"

```

```

repeat_region 14267..14759
/rpt_family="L1"
repeat_region 21435..21464
/rpt_family="(TA)n"
repeat_region 21691..21825
/rpt_family="L2"
repeat_region 23071..23094
/rpt_family="(TTTTG)n"
repeat_region 23232..23309
/rpt_family="CR1"
repeat_region 23572..23609
/rpt_family="AT_rich"
repeat_region 24164..24485
/rpt_family="Alu"
repeat_region 24282..24312
/rpt_family="(A)n"
repeat_region 24463..24494
/rpt_family="AT_rich"
repeat_region 25505..25717
/rpt_family="MaLR"
repeat_region 26100..26141
/rpt_family="AT_rich"
repeat_region 27299..27546
/rpt_family="L1"
repeat_region 27872..27694
/rpt_family="AT_rich"
repeat_region 28247..28485
/rpt_family="CR1"
repeat_region 29110..29183
/rpt_family="MaLR"
repeat_region 29874..29736
/rpt_family="CT-rich"
repeat_region 30236..30349
/rpt_family="L2"
repeat_region 30935..31360
/rpt_family="L1"
repeat_region 31847..31671
/rpt_family="AT_rich"
repeat_region 33127..33564
/rpt_family="ERV1"
repeat_region 33669..34235
/rpt_family="L1"
repeat_region 33871..33691
/rpt_family="AT_rich"
repeat_region 34236..34600
/rpt_family="MaLR"
repeat_region 34601..34721
/rpt_family="L1"
repeat_region 34724..35406
/rpt_family="L1"
repeat_region 35347..35368
/rpt_family="AT_rich"
repeat_region 35426..36225
/rpt_family="L1"
repeat_region 36288..38436

```

Query Match 7.6%; Score 21; DB 9; Length 155942;
 Best Local Similarity 100.0%; Pred. No. 0.86;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 TGTGTTTAAAGAAATCCTTTG 259

|||||

Db 71690 TGTGTTTAAAGAAATCCTTTG 71710

RESULT 14

AC096454

LOCUS

DEFINITION Rattus norvegicus clone CH230-4413, *** SEQUENCING IN PROGRESS ***;

ACCESSION AC096454

VERSION AC096454.4 GI:21723900

KEYWORDS HTG; HTGS_PHASE1.

164290 bp DNA linear HTG 11-JUL-2002

SOURCE
ORGANISM
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
1 (bases 1 to 164290)
Muzny,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Farraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlovic,J., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 164290)
Worley,K.C.

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 164290)
Worley,K.C.

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Jul 10, 2002 this sequence version replaced gi:20976243.

----- Genom Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFUG
Center clone name: CH230-44L3
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329

Consensus quality: 101666 bases at least Q40
Consensus quality: 108358 bases at least Q30
Consensus quality: 113449 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 73 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1066: contig of 1066 bp in length
1166: gap of unknown length
2367: contig of 1201 bp in length
2467: gap of unknown length
3860: contig of 1393 bp in length
3960: gap of unknown length
3961: contig of 1339 bp in length
5299: gap of unknown length
5300: contig of 1160 bp in length
6559: gap of unknown length
6659: gap of unknown length
7946: contig of 1287 bp in length
8046: gap of unknown length
9103: contig of 1057 bp in length
9203: gap of unknown length
10346: contig of 1143 bp in length
10446: gap of unknown length
11540: contig of 1094 bp in length
11640: gap of unknown length
13173: contig of 1533 bp in length
13273: gap of unknown length
14704: contig of 1431 bp in length
14804: gap of unknown length
15966: contig of 1162 bp in length
16066: gap of unknown length
17489: contig of 1423 bp in length
17589: gap of unknown length
17990: contig of 1199 bp in length
18788: gap of unknown length
18888: gap of unknown length
20000: contig of 1112 bp in length
20100: gap of unknown length
21279: contig of 1179 bp in length
21379: gap of unknown length
23233: contig of 1854 bp in length
23333: gap of unknown length
24363: contig of 1030 bp in length
24463: gap of unknown length
26083: contig of 1620 bp in length
26183: gap of unknown length
27576: contig of 1393 bp in length
27676: gap of unknown length
29028: contig of 1352 bp in length
29128: gap of unknown length
30635: contig of 1507 bp in length
30735: gap of unknown length
32888: contig of 2153 bp in length
32988: gap of unknown length
34437: contig of 1449 bp in length
34537: gap of unknown length
36387: contig of 1850 bp in length
36487: gap of unknown length
38380: contig of 1893 bp in length
38480: gap of unknown length
40384: contig of 1904 bp in length
40484: gap of unknown length
41907: contig of 1423 bp in length
42007: gap of unknown length
43661: contig of 1654 bp in length
43761: gap of unknown length
45748: contig of 1987 bp in length
43762

```

* 45749 45848: gap of unknown length
* 45849 47298: contig of 1450 bp in length
* 47299 47398: gap of unknown length
* 47399 48053: contig of 1455 bp in length
* 48054 50310: gap of unknown length
* 50311 52077: contig of 1357 bp in length
* 52078 52177: gap of unknown length
* 52178 53643: contig of 1667 bp in length
* 53644 53743: gap of unknown length
* 53744 55683: contig of 1466 bp in length
* 55684 55783: gap of unknown length
* 55784 57889: contig of 2106 bp in length
* 57890 60490: gap of unknown length
* 60491 60590: contig of 2501 bp in length
* 60591 62335: gap of unknown length
* 62336 62435: contig of 1745 bp in length
* 62436 64159: gap of unknown length
* 64160 64259: contig of 1724 bp in length
* 64260 65688: gap of unknown length
* 65689 65788: contig of 1429 bp in length
* 65789 67667: gap of unknown length
* 67668 69901: contig of 1879 bp in length
* 69902 70001: gap of unknown length
* 70002 71513: contig of 2134 bp in length
* 71514 71613: gap of unknown length
* 71614 73145: contig of 1512 bp in length
* 73146 73245: gap of unknown length
* 73246 75670: contig of 1532 bp in length
* 75671 75770: gap of unknown length
* 75771 77623: contig of 2425 bp in length
* 77624 77723: gap of unknown length
* 77724 80848: contig of 1853 bp in length
* 80849 82739: gap of unknown length
* 82740 82839: contig of 3025 bp in length
* 82840 85491: gap of unknown length
* 85492 85592: contig of 1891 bp in length
* 85593 88621: gap of unknown length
* 88622 88721: contig of 2852 bp in length
* 88722 91120: gap of unknown length
* 91121 91220: contig of 3030 bp in length
* 91221 93905: gap of unknown length
* 93906 94005: contig of 2399 bp in length
* 94006 97840: gap of unknown length
* 97840: contig of 2685 bp in length
* 97840: contig of 3835 bp in length

Query Match
Best Local Similarity 7.6%; Score 21; DB 2; Length 164290;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 AAGTATCAGGAGGAGGAGG 225
DB 139930 AAGTATCAGGAGGAGGAGG 139950
|||||
|||||

RESULT 15
AL359183
LOCUS
DEFINITION
AL359183
VERSION
AL359183.16 GI:16973851
KEYWORDS
HTC.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 180745)
AUTHORS
Tracey,A.
TITLE
Direct Submission

```

JOURNAL

COMMENT

Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Nov 17, 2001 this sequence version replaced g1:15131251.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at
<http://www.sanger.ac.uk/Projects/C.elegans/wormpep>
 This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>
 RP11-327E2 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-327E2 The true right end of clone RP11-5801 is at 37643 in this sequence.

FEATURES

source

```

1. 180745
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="10"
   /clone="RP11-327E2"
   /clone.lib="RPCI-11.2"
   60324..60396
   /note="Single clone region. Sequence from reads from a
   short insert library derived from a single pUC clone.
   Restriction digest data confirm the assembly."
   67015..67145
   /note="Single clone region. Sequence from reads from a
   short insert library derived from a single pUC clone.
   Restriction digest data confirm the assembly."
   84099..84252
   /note="Single clone region. Sequence from clone PCR only."
BASE COUNT 42272 a 44317 c 46166 g 47990 t
ORIGIN

```

Query Match

```

Best Local Similarity 7.6%; Score 21; DB 9; Length 180745;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 213 GGAGCAGGAGGATGGGAA 233

DB 54448 GGAGCAGGAGGATGGGAA 54468

Search completed: January 21, 2003, 13:50:52
 Job time : 1370.66 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: January 21, 2003, 09:57:51 ; Search time 379.797 Seconds
(without alignments)
4322.586 Million cell updates/sec

Title: US-10-060-830-3
Perfect score: 3888
Sequence: 1 MPFLILLVLLLEDAGA.....TQVSGAGRGCDVFKEL 729

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10060830/runat.16012003.091558.29002/app_query.fasta_1.1742
-DB=N.Geneseq.101002 -QWTF=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn.1.1.490.grunat.16012003.091558.29002 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DCPU=6 -ICPU=3
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3606	92.7	2046	21	AAZ51872 Human Factor 8 Hom
2	2806	72.2	2428	22	AAH34855 Human colon cancer
3	2592.5	66.7	1907	24	AAH35994 Human neuropilin-H
4	2592.5	66.7	1907	24	ABK49567 Human cDNA encodin
5	2592.5	66.7	1907	24	ABN59799 Novel human coding
6	2124.5	54.6	1388	24	ABR83640 Human cDNA differe
7	1056	27.2	3594	22	AAQ00660 Human TANGO 229 CD
8	1039	26.7	594	23	AAQ00660 DNA encoding novel
9	924.5	23.8	1620	22	AAQ00615 Novel human protei
10	924.5	23.8	1761	22	AAQ00614 Novel human protei
11	916.5	23.6	1768	22	AAQ00613 Novel human protei
12	897	23.1	640	21	AAQ80396 Human colon cancer
13	838	21.6	580	22	ABK60693 Human foetal liver
14	838	21.6	580	22	AAK08975 Human brain expres
15	838	21.6	580	22	AAK34866 Human bone marrow
16	838	21.6	580	22	AAI40582 Human foetal liver
17	828	21.3	467	22	ABA76819 Human brain expres
18	828	21.3	467	22	AAK25439 Human bone marrow
19	828	21.3	467	22	AAK51451 Probe #26227 used
20	828	21.3	467	22	AAI57541 Polynucleotide iso
21	574.5	14.8	1871	21	AAA96736 Human neuropilin-H
22	540.5	13.9	1265	24	AAH35992 Human cDNA encodin
23	540.5	13.9	1265	24	ABK49565 Human foetal liver
24	428	11.0	568	22	ABA60998 Human brain expres
25	428	11.0	568	22	AAK09290 Human bone marrow
26	428	11.0	568	22	AAK35179 Human genome-deriv
27	428	11.0	568	22	AAI40895 Rat semaphorin rec
28	428	11.0	568	24	ABSO9615 Mouse semaphorin r
29	426.5	11.0	3539	20	AAH08996 Mouse semaphorin r
30	425.5	10.9	4765	20	AAH09003 Mouse semaphorin r
31	425.5	10.9	4765	20	AAH09004 Mouse semaphorin r
32	423.5	10.9	4718	20	AAH08997 Mouse semaphorin r
33	423.5	10.9	4733	20	AAH08998 Mouse semaphorin r
34	423.5	10.9	4769	20	AAH08999 Mouse semaphorin r
35	423.5	10.9	4784	20	AAH09000 Mouse semaphorin r
36	423	10.9	3371	20	AAH08417 Neuropilin-2 gene
37	419.5	10.8	2670	21	AAH08417 Soluble neuropilli
38	418.5	10.8	6705	22	AAH32778 Human secreted pro
39	412	10.6	228	22	ABA73598 Human foetal liver
40	412	10.6	228	22	AAK22044 Human brain expres
41	412	10.6	228	22	AAK48207 Human bone marrow
42	412	10.6	228	22	AAI54037 Probe #22723 used
43	412	10.6	228	24	ABS22095 Human genome-deriv
44	411	10.6	2730	20	AAH09001 Human semaphorin r
45	411	10.6	2781	20	AAH09002 Human semaphorin r

ALIGNMENTS

RESULT 1
AAZ51872
ID AAZ51872 standard; cDNA; 2046 BP.
XX
AC AAZ51872;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human Factor 8 Homologue cDNA.
DE
KW Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW cerebroprotective; therapeutic; coagulation related disorder;
KW haemophilia; stroke; screening; ss.
XX
OS Homo sapiens.
XX
PH Key
FT CDS Location/Qualifiers
19..2028


```
QY 201 SerAsnThrLeuGlyGlyGlnIleSerValIleSerValIleSerLysGlyIleProTyrTyrGlu 220
DB 675 TCAACACGTTGGCGGCGCAATCAGTGTGTAATAGTAAGGTATYCCCTATTATGAA 734
QY 221 SerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr 240
DB 735 AGTCTTTGGGTACAACTGTCATCTGTGTGGGACACTTATCTACAGTCTTTTATACA 794
QY 241 PheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspPro 260
DB 795 TTTAGACAACTGGAGTGTATGGAACACTGGGATGGAGTCTGTGTGATCGCGATCCT 854
QY 261 GlnIleThrAlaSerSerValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrp 280
DB 855 CAATAACAGCATCATCTGTCTGGAGTGGACTGACCACACAGGCGCAAGAACAGATTGG 914
QY 281 LysProLysLysAlaArgLeuLysLysProGlyProGlyProTrpAlaAlaPheAlaThrAsp 300
DB 915 AAACCAAAAGCCAGGCTGAAACCTGGACCCCTTGGGTGCTTTTGGCCACTGAT 974
QY 301 GluTyrGlnTrpLeuGlnIleAspLeuAsnLysGlyLysIleThrGlyIleIleThr 320
DB 975 GAATACCACTGGTTACAAATAGATTGTAATAGGAAGAAATAACAGGCATTATAC 1034
QY 321 ThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSer 340
DB 1035 ACTGGAATCACCATTGGTGAGCACAATATCTATGTCTGTCTGCTCAGAAATCCTGTACAGT 1094
QY 341 AspAspGlyGlnLysTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePhe 360
DB 1095 GATGATGGCGAAGATGACGTGTGTACAGAGCGCTGGTGTGGAGCAAGATAAGATATT 1154
QY 361 GlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProProIleIleAla 380
DB 1155 CAAGAAACAAGATTATCACCAGGATGTGGTAAATCTTTTGGCCACCAATATTGCA 1214
QY 381 ArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeu 400
DB 1215 CGTTTATAGTAGTAATCCTACCAATGGCAGCAGCAAAATTTGGCATGAAATGGAGTGT 1274
QY 401 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProProPro 420
DB 1275 CTCGGATGTCAGTTTATTCCTAAAGTGTCTCTCAAAACTTACTCAACCTCCACCTCCT 1334
QY 421 ArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla 440
DB 1335 CGGAACACCAATGACCTCAAAACACTACAGCCCTCCAAAATAGCCAAAGTCTGTGCC 1394
QY 441 ProLysPheThrGlnProLysGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 460
DB 1395 CCAAAATTAGCGAACCACTACACCTGCAGTAGCAATGAATTTCTGTGCACAGACAGAA 1454
QY 461 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 480
DB 1455 CAACAACTGCCAGTCTGTATCAGAAATCTACCGTAATCTCAAAATGTAAACCAAGAT 1514
QY 481 ValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 500
DB 1515 GTAGCGCTGGCTGACGATCTTGTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCT 1574
QY 501 IleLeuValCysAlaThrPheTrpArgAsnArgLysLysLysThrGluGlyThrTyrAsp 520
DB 1575 ATATTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1634
QY 521 LeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMet 533
DB 1635 TTACCTTTACTGGGACCGGCGAGGTAACTCACTACGTGTGTCTT 1673
RESULT 3
AAD35994
ID AAD35994 standard; DNA; 1907 BP.
XX
```

```
AC AAD35994;
XX 09-AUG-2002 (first entry)
XX Human neuropilin-Hy2 DNA.
DE
XX
XX Human: neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW myotrophic-lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW myasthenia thyroditis; diabetes mellitus; graft-versus-host disease;
KW autohemne thyroiditis; autoimmune inflammatory eye disease; gene therapy;
KW neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive; chromosome 6q21; gene; ds.
XX
XX Homo sapiens.
OS
XX
XX Location/Qualifiers
PH Key 101..1258
FT CDS /*tag= a
FT /product= "Human neuropilin-Hy2 protein"
FT /note= "This region designated as SEQ.ID.NO:8 is
FT specifically referred in claim 1"
FT sig_peptide 101..160
FT /*tag= b
FT mat_peptide 161..1255
FT /*tag= c
FT /product= "Human mature neuropilin-Hy2 protein"
XX
XX WO200222815-A1.
XX
XX 21-MAR-2002.
XX
XX 12-SEP-2001; 2001WO-US28488.
XX
XX 11-SEP-2000; 2000US-0659671.
XX 06-SEP-2001; 2001US-317902P.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT;
XX
XX WPI; 2002-393966/42.
XX P-PSDB; AAE22716.
XX
XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
XX useful for treating neurodegenerative diseases e.g. Alzheimer's
XX disease, and for diagnosing and mapping genetic neuronal defects
XX
XX Claim 1; Page 126-128; 152pp; English.
XX
XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
XX neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
XX like polypeptides and polynucleotides are useful in modulating neuronal
XX growth regenerative capacity, treating neurodegenerative diseases,
XX diagnosing and mapping genetic neuronal defects and degenerative diseases
XX like Alzheimer's disease and for treating learning and memory disorders.
XX They are also useful for inducing angiogenesis, neovascularisation, as
XX well as organ growth and development e.g. heart and other tissues.
XX Antagonists of neuropilin-like polypeptides are useful for treating
XX cancers and other malignant diseases. Neuropilin is used to treat
XX platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
XX nocturnal haemoglobinuria and is used in nerve tissue growth or
XX regeneration, in wound healing, tissue repair and replacement and in
XX healing of bones, incisions and ulcers. Compositions comprising the
XX sequences of the invention are useful for treating diseases of peripheral
XX nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
```

CC syndrome, amytrophic lateral sclerosis, and traumatic disorders e.g.
CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
CC stroke, ulcers, immune deficiencies and immune disorders, infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, *Leishmania* spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy2
CC DNA. Neuropilin-Hy2 gene is located on chromosome 6q21.
XX
SO Sequence 1907 BP: 548 A: 449 C: 406 G: 504 T: 0 other:

QY 181 ProHisGlyTyrArgAspSerProLeuCyMetAlaGlyValHisAlaGlyValVal 200
DB 641 CCTCATGATATAGAGATCTCTCGCATTTGTCATGGCTGGTGTGTCATCGAGGAGTAGTG 700
QY 201 SerAsnThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGlu 220
DB 701 TCAAAACAGTGTGGCGGCCAAATCATGCTGTGTAATAGTAAAGGTATCCCTATTATGAA 760
QY 221 SerSerLeuAlaAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr 240
DB 761 AGTCTTGTGGCTAAACAGCTCACATCTGTGTGGGACACTTATACAGGTCTTTTATACA 820
QY 241 PheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspPr 260
DB 821 TTTAGACAGTGGATGTATGGACACTGGGATGGAGTCTGTGTGTGATCGGGATCC 880
QY 260 odinleThrAlaSerSerValLeuGluTyrThrAspHisThrGlyGlnGluAsnSerTr 280
DB 881 TCAAAATACAGCATCACT-GTGTGGAGTGGACTGACCCACACAGGGCAGAGAACAGTTG 939
QY 280 pLys-ProLysLysAlaGlyLeuLysLysProGlyProProTrpAlaAlaPheAlaThrA 300
DB 940 GAAACCCAAAACAGGCTGAAAACCTGGACCCCTGGGT-GCTTTTGCCACTG 998
QY 300 SPGLuTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleT 320
DB 999 ATGAATACAGTGTGTACAAATAGATTGTAATAGGAAAGAAAATAACAGGCATTATAA 1058
QY 320 hrThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrS 340
DB 1059 CCACGTGGATCCACCATGGT-GAGCACAATTACTATGTCTGCTGTACAGAACTCCTGTACA 1117
QY 340 erAspAspGlyGlnLysTrpThrValTyrArgGluProGly-ValGluGlnAspLysIle 359
DB 1118 GTGATGTGGCGGAGAAATGGAGCTGTACAGAGACCTGGCTGGGACACAGATAGATA 1177
QY 360 PheGlnGlyAsn--LysAspTyrHisGlnAsp-ValArgAsnAsnPheLeuProPhe 378
DB 1178 TTTCAAGGAAACAAAAGAAATTAACAGGATGGTGTGTAATTAATTTTGGCACCAATT 1237
QY 379 IleAlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMet 398
DB 1238 ATTGCAGC-TTTATTAGAGTGAATCCTACCCAAATGGCAGCAAAATTTGCCATGAAATG 1296
QY 399 GluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProPro 418
DB 1297 GAGCTGCTGGATGTCAGTTTATTCTTAAGGTCTGCTCCAAACTTACTCAACCTCCA 1356
QY 419 ProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLys---IleAlaLys 437
DB 1357 CCTCTCGGAACAGCAATGACCTCAA-AACACTACAGCCCTCCCAAAAATTTAGCCAAA 1415
QY 438 GlyArgAlaProLysPheThrGlnProLysGlnProArgSerSerAsnGluPheProAla 457
DB 1416 GTCTGTGCCCCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATGAATTTCTCTCA 1475
QY 458 GlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnVal 477
DB 1476 CAGACAGAACAACTCCAGTCTGTATATCAGAAATATACCGTAACTCCAATGTA 1535
QY 478 ThrLysAspValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThr 497
DB 1536 ACCAAGATGTAGCGCTGCATGCTTGTCTCCCTGTGCTGTGCTCATGCTCTCACTACT 1595
QY 498 LeuIleLeuLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGluGly 517
DB 1596 CTCAATCTCATATTAGTGTGCTTGGCAGCTGGAGAAACAGAGAAAGAAAACCTGAAGGC 1655
QY 518 ThrTyrAspLeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMet 533
DB 1656 ACCTATGACTTACCTTACTGGGACCGGCGAGGTAACCTACAGTGGTCTT 1703

RESULT 6
ABK83640
ID ABK83640 standard; cDNA; 1388 BP.
XX
AC ABK83640;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #211.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
XX 03-OCT-2001; 2001WO-US30821.
XX
XX 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
XX WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity
XX
PS Claim 1; SEQ ID No 211; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present

CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1388 BP; 432 A; 289 C; 302 G; 364 T; 1 other;

Alignment Scores:

Pred. No.: 2,76e-151 Length: 1388
 Score: 2124.50 Matches: 442
 Percent Similarity: 95.10% Conservativeness: 4
 Best Local Similarity: 94.24% Mismatches: 11
 Query Match: 54.64% Indels: 17
 DB: 24 Gaps: 3

US-10-060-830-3 (1-729) x ABK83640 (1-1388)

Qy 24 AspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerIleAsnTyr 43
 Db 3 GATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAAACCTTACATCCATAAATAC 62
 Qy 44 ProGlnThrTyrProAsnSerThrValCysGluTrpGluIleArgValLysMetGlyGlu 63
 Db 63 CCACAGACCTATCCCAACAGACACTGTTTGTGAATGGGAGATCCGTTAAAGATGGGAGAG 122
 Qy 64 ArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCysHisPheAsn 83
 Db 123 AGAGTTCGCATCAAAATTTGGTGACATTTGACATTTGAAGATTCGATTCCTGTGCATTTAT 182
 Qy 84 TyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLysTyrCysGly 103
 Db 183 TACTTGAGAAATTTAATAGGAATGGAGTCAGCAACACTGAAATAGGCAAACTACTGTGT 242
 Qy 104 LeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPhe 123
 Db 243 CTGGGGTTGGCAATCAACCATCAATGAATCAAAAGGCAATGAATCAATCATCTGTGTTC 302
 Qy 124 MetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLys 143
 Db 303 ATGAGTGGAAATCCATGTTCTGGACGGGATTTTGGCCTCATCTCTGTTATAGATATA 362
 Qy 144 Gln-AspLeuIleThrCysLeuAspThrAlaSerAsn-PheLeuGluProGluPheSerL 163
 Db 363 CAAGATCTAATTAATCTGTTGGACACTGCATCCCAATTTTGGAACTGAGTTCAGTA 422
 Qy 163 ySTyrCys-ProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIleProHis 182
 Db 423 AGTACTGCCCCAGCTGGTGTCTGCTTCTCTGCTGAGATATCTGGAACAATTCCTCAT 482
 Qy 183 GlyTyrArgAspSerProLeuCysMetAlaGlyValHisAlaGlyValValSerAsn 202
 Db 483 GGATATAGAGATTCCTGCCCATTTGTCATGGCTGGTGTGCATGAGGAGTAGTGTCAAC 542
 Qy 203 ThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGluSerSer 222
 Db 543 ACCTTTGGCGGCCAAATCAGTCTTCTAATAGTAAAGGATTCCTTATATGAAAGTTCT 602
 Qy 223 LeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeu---PheThrPhe 241
 Db 603 TTGGGTAAACAGCTCACATCTCTGTGGGACACTTGATACACTCAAGTCTTTTTTACATTT 662
 Qy 242 LysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGly-ValIleAlaAspProGlu 261
 Db 663 AAGACAAGTGGATTTATGGAACACTGGGNATGGAGTCTGGTGGTGATCCGGATCCTCA 722
 Qy 261 nIleThrAlaSerSerValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrpLy 281
 Db 723 AATAACAGCATCACT-GTCTCTGAGTGGACTGACCACACAGGCGCAAGAACAGATTGGAA 781
 Qy 281 s-ProLysLysAlaArgLeuLysLysProGlyProProTrpAlaAlaPheAlaThrAspG 301
 Db 782 ACCCAAAAAAGCCAGGCTGAAAAAACCTGGACCGCTTGGGT-GCTTTTGGCACTGATG 840

Qy 301 LuTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThr 321
 Db 841 AATACCACTGCTTACAAATAGATTGATAAGGAAAAAGAAATAACAGGCATTTATAACCA 900
 Qy 321 hrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerA 341
 Db 901 CTGGATCCACCATGGT-GAGCACAAATTACTATGTCTGTGCTCAGCAATCCTGTACAGTG 959
 Qy 341 sPaspGlyGlnLysTrpThrValTyrArgGluProGly-ValGluGlnAspLysIlePhe 360
 Db 960 ATGATGGCCAGAAATGGACTGTGTACAGAGAGCCCTGGTGGAGCAAGATAAGATATTT 1019
 Qy 361 GlnGlyAsn--LysAspTyrHisGlnAsp-ValArgAsnAsnPheLeuProProIleIle 379
 Db 1020 CAAGGAACAAAGAAATATCACCAGGATGTGTGCGTAATAACTTTTCCCACTTATTT 1079
 Qy 380 AlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGlu 399
 Db 1080 GCACG-TTTATTAGAGTGAATCTTACCAATGGCAGCAGAAAAATGGCATGAAATGGAG 1138
 Qy 400 LeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProPro 419
 Db 1139 CTGCTCGGATGTCACTTATTCTTAAAGTGTCTCTCAAAACTTACTCACTCCACCT 1198
 Qy 420 ProArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLys---IleAlaLysGly 438
 Db 1199 CCTCGGAACAGCAATGACCTCAA-AACACTACAGCCCTCCAAAAAATTTAGCCAAAGT 1257
 Qy 439 ArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGln 458
 Db 1258 CGTGCCCCAAAAATTTACGCAACCACTACAACTCGCAGTAGCAATGAATTTCTCTCAG 1317
 Qy 459 ThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrVal---ThrProAsnVal 477
 Db 1318 ACAGAACCAACCACTGCCAGTCTCTGATATCAAAAAATTTACTTACCGTAAACTCCAAATGTA 1377
 Qy 478 ThrLysAsp 480
 Db 1378 ACCAAAGAT 1386

RESULT 7
 AAS00660
 ID AAS00660 standard; cDNA; 3594 BP.
 XX AC AAS00660;
 XX DT 07-SEP-2001 (first entry)
 XX DE Human TANGO 229 cDNA.

Human: TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
 skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
 bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
 anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
 malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
 attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
 hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
 muscular dystrophy; immuno-competence; vertebrae; blood; serum.

Homo sapiens.

Key	Location/Qualifiers
PH	72..2219
FT	CDS
FT	/tag= a
FT	/product= "Human TANGO 229"
FT	72..173
FT	/tag= b
FT	174..2216
FT	/tag= c
FT	/product= "Mature human TANGO 229"
XX	PN WO200129088-A1.


```
Qy 477 ValThrLysAspValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThr 496
Db 1446 -----GTGGCTATCCATGGTCTCTCTGTCCTGGTGGTTGCTGGAATG 1493
Qy 497 ThrLeuLeuLeuLeuValCysAlaThrPHistrprgAsnArgLysLysLysThrGlu 516
Db 1494 GGGATCTTTCAGCGCTTT-----AGAAAAGAGAAGAAGAAA 1529
Qy 517 GlyThrTyrAspLeuProTyrTrp-----AspArgAlaGlyTriTyrLysGly 532
Db 1530 GGAAGT-----CCGTATGGATCAGACAGCGCTCAGAAAACAGACTGTTGGAAGCAG 1580
Qy 533 MetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSer 552
Db 1581 ATTAATATCCCTTT-----GCCAGACATCAG 1607
Qy 553 SerSerGlu-----ValAsnHisLeuSerProArgGluValThr-----ThrVal 567
Db 1608 TCAGCTGAGTTTACCATCAGCTATGATGATGAGAAGAGAGATGACACAAAAGTTAGATCTC 1667
Qy 568 LeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeu 587
Db 1668 ATCAAGTATGATGGCAGATTACACAGCCCTCTATGATTGGCAGCGGACAGTCCAG 1727
Qy 588 HisGlnArgSerThrPheLysPro-----GluGluGlyLysGluAlaGlyTyrAlaAsp 605
Db 1728 AGGAAGGCTCCACCTTCGGCCCATGGACACGGATGCGGAGGAGGCGAGG---GTGAGC 1784
Qy 506 LeuAspPro-----TyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGlu 622
Db 1785 ACCATCGCGCGCCCATATGACTGCGCCGAGCGCGCGCCCGCCACGAGTACGCGGTG 1844
Qy 623 ProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetAspMetSerGlyHis 642
Db 1845 CCCTGCGCGCGCGCGAGCGGATGCGCCACGCGCCATCGTG-----GAGCGGCGAC 1895
Qy 643 ProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnPro 662
Db 1896 GTGCTGCGCGCGCCACAGTCTCTCTGGGAGAGCGGCTACCGCGTCCCGCGCCCGAGCC 1955
Qy 663 -----ProProLeu-----ValGlyThrTyr 669
Db 1956 GGCACAAACACTCCCTCTCTCGGGCGGTCTCTCCCGGTAGCGGGTGGGCGCCCGAC 2015
Qy 670 AsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrPro 689
Db 2016 GACGAGACTATCAAGGCCACACAGCCACAGCGCTGCGGACAGGGGCTACGACCGGCC 2075
Qy 690 LysAla-GlyLysProGlyLeuProAlaProAspGluLeuValTyrGlnValProGln 708
Db 2076 AAAGCTGTACGGCGCTCGCCACCGAAGCGGACACCCCTGACTCTCAGAAGCCGCCAA 2133
RESULT 8
AAS65485
ID AAS65485 standard; cDNA; 594 BP.
XX AC AAS65485;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #1289.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
```

```
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG01298.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 1; SEQ ID No 1289; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS6197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 594 BP; 175 A; 160 C; 139 G; 120 T; 0 other;

Alignment Scores:
Pred. No.: 9.73e-70 Length: 594
Score: 1039.00 Matches: 197
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.72% Indels: 0
DB: Gaps: 0

US-10-060-830-3 (1-729) x AAS65485 (1-594)
Qy 533 MetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSer 552
Db 1 ATGAGCAGCTTCTCTGCAAAAGCAGTGGACCATGAGGAACCCAGTTCGCTATAGC 60
Qy 553 SerSerGluValAsnHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSer 572
Db 61 AGCAGCGAAGTTAATCACTGAGTCCAAAGAGAAGTCACACAGCTGCTCAGGCTGACTCT 120
Qy 573 AlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThr 592
Db 121 GCAGAGTATGCTCAGCCCATGTTAGGAGGAATGTTGTGTACACTTCATCAAGATCTATCC 180
Qy 593 PheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerPro 612
Db 181 TTTAAACCCAGAGAAGAAAAGAGACAGCGCTATGCAGACCTAGATCCTTACACTCACCA 240
Qy 613 GlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAla 632
Db 241 GGGCAGGAAGTTTATCATGCTCTGCTGAACCACTCCCAATACGGGCGCTGAGTATGCA 300
Qy 633 ThrProIleIleMetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThr 652
```


XX WPI; 2001-290917/30.
 DR P-PSDB; AAU00628.
 XX Novel nucleic acid encoding human CUB-domain containing protein, useful
 PT for drug screening, diagnosis and treatment of physiological disorders
 PT or diseases -
 XX Disclosure; Page 30-31; 33pp; English.
 PS
 XX The sequence represents a polynucleotide which encodes a novel human
 CC protein (NHP) containing a CUB domain (an extracellular domain). CUB
 CC proteins have been associated with regulating development, modulating
 CC cellular processes and preventing infectious disease. NHP nucleotide
 CC sequences are useful for gene therapy of physiological disorders or
 CC diseases. NHP oligonucleotides are useful as hybridisation probes for
 CC screening libraries and assessing gene patterns. NHP nucleotide sequences
 CC are useful for detecting mutant or inappropriately expressed NHPs (for
 CC example, those proteins associated with obesity, high blood pressure,
 CC connective tissue disorders and infertility) for the diagnosis of a
 CC disease. The polynucleotides may also be used in screening for drugs
 CC effective in the treatment of symptomatic or phenotypic manifestations of
 CC perturbing the normal function of NHP in the body. Nucleotide constructs
 CC encoding NHP products are used to genetically engineer host cells to
 CC express such products in vivo. These host cells allow for the
 CC identification of compounds that bind to NHP receptors or trigger
 CC NHP-mediated pathways.
 XX
 SQ Sequence 1768 BP; 539 A; 370 C; 421 G; 438 T; 0 other;

Alignment Scores:
 Pred. No.: 7,04e-60 Length: 1768
 Score: 916.50 Matches: 215
 Percent Similarity: 54.75% Conservative: 96
 Best Local Similarity: 37.85% Mismatches: 176
 Query Match: 23.57% Indels: 81
 DB: 22 Gaps: 16

US-10-060-830-3 (1-729) x AAS00613 (1-1768)

Qy 23 GlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSertileAsn 42
 Db 14 GGTATGGCTGTGGACACCTAGTACTATCAGATAGTAGGACCAATACATCTAAGAAAT 73
 Qy 43 TyrProGlnThrThrProAsnSerThrValCysGluThrPpGluileArgValysMetGly 62
 Db 74 TATCCGGGACCTACCCCAATCACACTGTTTCCGAAAGACAAATTACAGTACCAGGGG 133
 Qy 63 GluArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCysHisPhe 82
 Db 134 AAAAGACTGATCTGAGGTGGAGATTGGATATCGAA--TCCAGACCTGTGCTTCT 190
 Qy 83 AsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLysTyrCys 102
 Db 191 GACTATCTTCTTCACACGC-----TCTTCAGATCAATATGGTCCATCTGT 238
 Qy 103 GlyLeuGlyLeuGlnMetAsnHisSertileGluSerLysGlyAsnGluIleThrLeuLeu 122
 Db 239 GGA---AGTATGACTGTTCCTCCCAAGAACTCTGTTTGAACACAAAGTGAAGTACCCTCGC 295
 Qy 123 PheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValleAsp 142
 Db 296 TTTGAGATGGATCCCATTTCTGGCCGGGGTTTTTGTGACCTATCGGACGACGCAC 355
 Qy 143 LysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPheSer 162
 Db 356 CATCAGATTTAATAACATGTTTGAACAGCAGCTAGCCATTATTTGAAGACAAATACACG 415
 Qy 163 LysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIleProHis 182
 Db 416 AAATTCCTCCAGCTGGTGTAGACACGTAGCAGGACGACATTTCTGGGAATATGTAGAT 475
 Qy 183 GlyTyrArgAspSerProLeuCysMetAlaGlyValHisAlaGlyValValSerAsn 202

Db 476 GGATATAGACATACCTCTTTATTGTGCAAGCTGCCATCCATCGAAGTAATTTGCTGAT 535
 Qy 203 ThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGluSerSer 222
 Db 536 GAACCTAGGTGGCCAGATCAGTGTCTTCAGCGCAAGGATCAGTCGATATGAAGGATT 595
 Qy 223 LeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThrPheLys 242
 Db 596 CTGGCCAAATGGTGTCTTTCGAGGGATGGTCCCTTCAGACAAAGCATTGTGTTTACC 655
 Qy 243 ThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIle 262
 Db 656 TCCAAATGGTTCAGCAGCATCTGTTGAGTTTGA-----CCTGACGGCAATC 703
 Qy 263 ThrAlaSerSerValLeuGluThrThrAspHisThrGlyGlnGluAsnSerTyrPlyPro 282
 Db 704 AGAGCTTCTTCTCATGCGCAGTCGTCATGAGAGTGGAGACCAAGTTCACCTGCTCCT 763
 Qy 283 LysLysAlaArgLeuLysLysProGlyProThrAlaAlaPheAlaThrAspGlu--- 301
 Db 764 GGCCAAAGCCGACCTCAGGACCAAGGCCCATCATGGGCTTCGGGCGCAGCAGTAGCAACAAC 823
 Qy 302 -----TyrGlnThrLeuGlnIleAspLeuAsnLysGlyLysIleThrGlyIle 318
 Db 824 CACAAACCCAGCAGAGTGGCTGGAGATCGATTTGGGGAGAGAAAAGAAATAACAGGAAT 883
 Qy 319 IleThrThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeu 338
 Db 884 AGGACCACAGGATCTACACAGTCGAACTTCAACTTTTATGTTAAGAGTTTGTGATGAAC 943
 Qy 339 TyrSerAspAspGlyGlnLysThrThrValTyrArgGluProGlyValGluGlnAspLys 358
 Db 944 TTCAAAAACAAATAATTTCAAGTGAAGACCTATAAGGAATTTGAATAATTAAGAAAG 1003
 Qy 359 IlePheGlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProPhe 378
 Db 1004 GTGTTTCAGGTAACCTCAACTTTCGGGACCCAGTGGCAAAACAATTTATCCCTCCCATC 1063
 Qy 379 IleAlaArgPheIleArgValAsnProThrGlnThrGlnGlnLysIleAlaMetLysMet 398
 Db 1064 GTGGCCAGATATGTGCGGTTGTCCCCAGACATGGCCAGCAGAGTAGCCTTGAAGGTG 1123
 Qy 399 GluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnPro 418
 Db 1124 GAGCTCATTTGTTGCCAG-----ATTACACA----- 1150
 Qy 419 ProProArgAsnSerAsnAsp-----LeuLysAsnThrThrAlaProProLys 434
 Db 1151 -----GGTAATGATTCATGTTGGTGGCCCAAGACAAAGTCAAGCACCAGCTGTT 1198
 Qy 435 IleAlaLysGlyArgAlaProLysPheThrGlnProLysGlnProArgSerSerAsnGlu 454
 Db 1199 TCAACTAAGAAGAAGATGAGACAAATCACAAGGCCATC----- 1237
 Qy 455 PheProAlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThr 474
 Db 1238 -----CCCTCGGAAGAACA-----TCCAGAGAAATAAACATTACAACG----- 1276
 Qy 475 ProAsnValThrLysAspValAlaLeuAlaValLeuValProValLeuValMetVal 494
 Db 1277 -----GTGGCTATTCATGTTGGTGGCTTCTGCTGCTGCTGCTGCTGCTGCT 1318
 Qy 495 LeuThrThrLeuIleLeuValCysAlaThrPheHisTrpArgAsnArgLysLysLys 514
 Db 1319 GGAATGGGATCTTTCAGCCCTTT-----AGAAAGAAGAAG 1354
 Qy 515 ThrGluGlyThrThrAspLeuProTyrTrp-----AspArgAlaGlyTrpTrp 530
 Db 1355 AAGAAGGAAGT-----CCGTATGATCAGCGGAGCTCAGAAAACACACTGTTGG 1405
 Qy 531 LysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProValArg 550
 Db 531 LysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProValArg 550

Db 1406 AACGAGTAAATATCCCTTT-----GCCAGA 1432
QY 551 TyrSerSerSerGlu-----ValAsnHisLeuSerProArgGluValThrThrValLeu 568
Db 1433 CATCAGTCAGCTGAGTTTACCATCAGCTATGATGATGAGAGGAGATGACACAAAGTTA 1492
QY 569 GlnAlaAspSerAlaGluTyrAla 576
Db 1493 GATCTCATCACAAGTGATGGCA 1516
RESULT 12
AAZ80396/C
ID AAZ80396 standard; cDNA; 640 BP.
XX
AC AAZ80396;
XX
DT 07-APR-2000 (first entry)
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:480.
XX
KW Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytosolic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO9964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB) BAYER CORP.
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI; 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer -
XX
PS Claim 15; Page 322; 469pp; English.
XX
CC AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC can be used for developing agents for the diagnosis and treatment of
CC disorders involving unwanted cell proliferation, such as neoplasia,
CC dysplasia or hyperplasia.
XX
SQ Sequence 640 BP; 148 A; 134 C; 160 G; 192 T; 6 other;

Alignment Scores:
Pred. No.: 5.74e-59 Length: 640
Score: 897.00 Matches: 194
Percent Similarity: 91.98% Conservative: 1
Best Local Similarity: 91.51% Mismatches: 17
Query Match: 23.07% Indels: 6
DB: 21 Gaps: 0

US-10-060-830-3 (1-729) x AAZ80396 (1-640)

QY 375 LeuProPheIleAlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIle 394

Db 630 TTGCCACCAATATGGCAGC-TTTATTAGAGT-AATCCTTACCCATGGCAGCAAAATT- 574
QY 395 AlaMetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProLysLeu 414
Db 573 GCCATGAAAT-GAATGCTCGGATGTCAGTTTATTCTAAAGT-NGTCNTCAAACCTT 516
QY 415 ThrGlnProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProLys 434
Db 515 ACTCAACCTCCACCTCTTNGNAACAGCAATGNCCTCAAAACACCTACAGCCCTCCAAA 456
QY 435 IleAlaLysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGlu 454
Db 455 TA-GCCAAAGGTCGTGCCCAAAATTTACGCAACACCTACAACTTCGCAGTAGCAATGAA 397
QY 455 PheProAlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThr 474
Db 396 TTTCTCTCACAGACAGACAAACAACTGCCAGTCTGTATATCAGAAATACCTACCCGTA 337
QY 475 ProAsnValThrLysAspValAlaLeuAlaValLeuValProValLeuValMetVal 494
Db 336 CCAAATGTAACCAAGATGATGCGCTGCTGCAGTTCTTGTCCCTGCTGCTCATGGTC 277
QY 495 LeuThrThrIleLeuLeuValCysAlaTrpHisTrpArgAsnArgLysLys 514
Db 276 CTCACCTACTCTCATTTCTCATATTAGTGTGCTTGGCACTGGAGAAACAGAAAGAAA 217
QY 515 ThrGluGlyThrTyrAspLeuProTyrTrpAspArgAlaGlyTyrTrpLysGlyMetLys 534
Db 216 ACTGAAGCCACTATGACTTACCTTACTGGGGCCGGCAGGTGGTGGAAAGGAATGAG 157
QY 535 GlnPheLeuProAlaLysAlaValAspHisGluThrProValArgTyrSerSerSer 554
Db 156 CAGTTTCTCTGCAAAAGCAGTGGACCATGAGGAACCCAGTTCGCTATAGCAGCAG 97
QY 555 GluValAsnHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGlu 574
Db 96 GAATTAATACCTACCTGAGTCCAGAGAGAGTCCACACAGTGTGCTGAGGCTGACTCTGCAGAG 37
QY 575 TyrAlaGlnProLeuValGlyGlyIleValGlyThr 586
Db 36 TATGCTCAGCCACTGGTAGGAGGAATTTGTTGTACC 1

RESULT 13
ABA60693 standard; DNA; 580 BP.
ID ABA60693
XX
AC ABA60693;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #8998.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-483447/52.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX Claim 1; SEQ ID NO 8998; 639pp + sequence listing; English.
 PS The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 1.45e-54 Length: 580
 Score: 838.00 Matches: 161
 Percent Similarity: 95.38% Conservative: 4
 Best Local Similarity: 93.06% Mismatches: 7
 Query Match: 21.55% Indels: 1
 Gaps: 0
 DB:
 US-10-060-830-3 (1-729) x ABA60693 (1-580)
 QY 558 HisLeuSerProArgGluValThrValLeuGlnAlaAsp-SerAlaGluTyrAlaG1 577
 DB 21 AATCTCTTACTAGTAAATGACCTTTCTGTTGTTGATATCCACAGAGTATGCTCA 80
 QY 577 nProLeuValGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGlu1 597
 DB 81 GCCACTGGTAGGAGGAATGTGGTACACTTCATCAAGATCTACTTTAAACCCAGAGA 140
 QY 597 uGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTy 617
 DB 141 AGGAAAGACAGCGCTATGACAGCTAGATCCTTACACTCACCAGGCGAGGAGTTTA 200
 QY 617 rHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMe 637
 DB 201 TCATGCTATGCTGAACCACTCCCAATACGGGGCTGAGTATGCAACCCCAATCATCAT 260
 QY 637 tAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheIysAl 657
 DB 261 GGACATGTCCAGGGCACCACCACTTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGC 320
 QY 657 aThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAs 677
 DB 321 TACGGGGACCAACCTCCCACTAGTGGGACTTACATACACTCTCTCCAGCAGTCA 380
 QY 677 pSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuPr 697
 DB 381 CAGCTGCTCCTCAGCCAGGCCAGTATGATATACCCGAAAGCTGGGAAAGCCAGGCTACC 440
 QY 697 oAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaG1 717
 DB 441 TGCCCCACACGCAATTTGGTGTACAGGTGCCAGAGACACACAGAAGTATCAGGAGCAGG 500
 QY 717 yArgAspGlyGluCysAspValPheLysGluIleLeu 729
 DB 501 AAGGGATGGGAATGTGATGTTTTTAAGAATTCCTT 537
 RESULT 14
 ID AAK08975
 XX AAK08975 standard; DNA; 580 BP.
 AC AAK08975;
 XX

DT 05-NOV-2001 (first entry)
 XX Human brain expressed single exon probe SEQ ID NO: 8966.
 XX Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.
 XX Homo sapiens.
 OS WO200157275-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00667.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PT WPI; 2001-483446/52.
 DR Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX Example 4; SEQ ID NO: 8966; 650pp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.
 XX Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 1.45e-54 Length: 580
 Score: 838.00 Matches: 161
 Percent Similarity: 95.38% Conservative: 4
 Best Local Similarity: 93.06% Mismatches: 7
 Query Match: 21.55% Indels: 1
 Gaps: 0
 DB:
 US-10-060-830-3 (1-729) x AAK08975 (1-580)
 QY 558 HisLeuSerProArgGluValThrValLeuGlnAlaAsp-SerAlaGluTyrAlaG1 577
 DB 21 AATCTCTTACTAGTAAATGACCTTTCTGTTGTTGATATCCACAGAGTATGCTCA 80
 QY 577 nProLeuValGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGlu1 597
 DB 81 GCCACTGGTAGGAGGAATGTGGTACACTTCATCAAGATCTACTTTAAACCCAGAGA 140
 QY 597 uGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTy 617
 DB 141 AGGAAAGACAGCGCTATGACAGCTAGATCCTTACAACTCACCAGGCGAGGAGTTTA 200
 QY 617 rHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMe 637
 DB 201 TCATGCTATGCTGAACCACTCCCAATACGGGGCTGAGTATGCAACCCCAATCATCAT 260
 QY 637 tAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheIysAl 657
 DB 261 GGACATGTCCAGGGCACCACCACTTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGC 320
 QY 657 aThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAs 677
 DB 321 TACGGGGACCAACCTCCCACTAGTGGGACTTACATACACTCTCTCCAGCAGTCA 380
 QY 677 pSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuPr 697
 DB 381 CAGCTGCTCCTCAGCCAGGCCAGTATGATATACCCGAAAGCTGGGAAAGCCAGGCTACC 440
 QY 697 oAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaG1 717
 DB 441 TGCCCCACACGCAATTTGGTGTACAGGTGCCAGAGACACACAGAAGTATCAGGAGCAGG 500
 QY 717 yArgAspGlyGluCysAspValPheLysGluIleLeu 729
 DB 501 AAGGGATGGGAATGTGATGTTTTTAAGAATTCCTT 537
 RESULT 14
 ID AAK08975
 XX AAK08975 standard; DNA; 580 BP.
 AC AAK08975;
 XX

Db 261 GGACATGTGAGGCGACCCCAACATTCAGTTGGTCAGCGCTCCACATCCACTTTCAAGGC 320
 QY 657 aThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAs 677
 Db 321 TACGGGGAAACCAACTCCCCCACTAGTGGAACTTACAACTACACATCTCTCCAGGACTGA 380
 QY 677 pSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuPr 697
 Db 381 CAGCTGCTCCTCAGCCAGCGCCAGTATGATATACCCGAAAGCTGGGAAGCCAGGTCTACC 440
 QY 697 oAlaProaspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaG1 717
 Db 441 TGCCCCAGACGAATTGGTGTACCAAGTCCACAGACGACACAGAAGATATCAGGAGCAGG 500
 QY 717 YArgAspGlyGluCysAspValPheLysGluLeuLeu 729
 Db 501 AAGGGATGGGAATGTGATGTTTTTAAAGAAATCCTT 537

RESULT 15
 AAK34866
 ID AAK34866 standard; DNA; 580 BP.
 XX
 AC AAK34866;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 9423.
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0623366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SS, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 9423; 658pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX
 SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;

Alignment Scores:
 Pred. No.: 1.45e-54 Length: 580
 Score: 838.00 Matches: 161
 Percent Similarity: 93.38% Conservative: 4
 Best Local Similarity: 93.06% Mismatches: 7
 Query Match: 21.55% Indels: 1

DB: 22 Gaps: 0
 US-10-060-830-3 (1-729) x AAK34866 (1-580)
 QY 558 HisLeuSerProArgGluValThrThrValLeuGlnAlaAsp-SerAlaGluTyrAlaG1 577
 Db 21 AATCTCTTTACTAGTAAATGACATTTCTGTGTTGTTTGTATATCCACAGAGTATGCTCA 80
 QY 577 nProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluG1 597
 Db 81 GCCACTGTAGGAGGAATTTGTTGTACACTTCATCAAAAGATCTACCTTTAAACCAAGA 140
 QY 597 uGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTy 617
 Db 141 AGGAAAGAGCAGCGCTATGCAGACTTAGATCCTTACAACTCACCGAGGCGAGGAGTTTA 200
 QY 617 rHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMe 637
 Db 201 TCATGCTATGCTGAACCACTCCCAATTTACGGGCGCTGAGTATGCAACCCCAATCATCAT 260
 QY 637 tAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAl 657
 Db 261 GGACATGTGAGGCGACCCCACTTCACTAGTTGGTCAGCCCTCCACATCCACTTTCAAGGC 320
 QY 657 aThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAs 677
 Db 321 TACGGGGAACCACTCCCCCACTAGTGGAACTTACAACTACTTCTCTCCAGGACTGA 380
 QY 677 pSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuPr 697
 Db 381 CAGCTGCTCCTCAGCCAGCGCCAGTATGATATACCCGAAAGCTGGGAAGCCAGGTCTACC 440
 QY 697 oAlaProaspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaG1 717
 Db 441 TGCCCCAGACGAATTGGTGTACCAAGTCCACAGACGACACAGAAGATATCAGGAGCAGG 500
 QY 717 YArgAspGlyGluCysAspValPheLysGluLeuLeu 729
 Db 501 AAGGGATGGGAATGTGATGTTTTTAAAGAAATCCTT 537

Search completed: January 21, 2003, 10:13:09

Job time : 396.797 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:53:55 ; Search time 28.4848 Seconds
(without alignments)
3410.229 Million cell updates/sec

Title: US-10-060-830-3

Perfect score: 729

Sequence: 1 MFLFLLLVLLVLLLEDAGA.....TQEVGAGRGDCVDFKEIL 729

Scoring table:

OLIGO

Gapop 60.0 , Capext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_101002.*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	669	91.8	669	21	Human Factor 8 Hom
2	326	44.7	583	22	Human colon cancer
3	255	35.0	385	23	Human neuropilin-H
4	255	35.0	385	23	Human Neuropilin-H
5	255	35.0	385	23	Novel human protei
6	235	32.2	365	23	Human neuropilin-H
7	197	27.0	197	22	Novel human diagno
8	75	10.3	75	22	Peptide #9723 enco
9	75	10.3	75	22	Human brain expres
10	75	10.3	75	22	Human bone marrow

11	75	10.3	75	22	AA136024	Peptide #10061 enc
12	75	10.3	75	23	ABG45349	Human peptide enco
13	20	2.7	20	23	AAE22720	Human neuropilin-H
14	10	1.4	629	23	AAU74424	Mouse protein sequ
15	10	1.4	677	19	AAW70900	Human beta galacto
16	10	1.4	696	21	AB56675	Human prostate can
17	9	1.2	73	22	AA62673	Propionibacterium
18	9	1.2	89	23	ABG64606	Human albumin fusi
19	9	1.2	89	23	AAE21629	Human gene 1 encod
20	9	1.2	150	23	AAU83189	Novel secreted pro
21	9	1.2	150	23	AAU50708	Human secreted cyt
22	9	1.2	209	23	AAU83137	Novel secreted pro
23	9	1.2	276	19	AAW98563	H. pylori GHPO 174
24	9	1.2	377	22	AAU38998	Human polypeptide
25	9	1.2	377	23	AAE14752	Human WKL-1 polype
26	9	1.2	441	22	AAU41041	Human polypeptide
27	9	1.2	700	22	AAU23827	Human EST encoded
28	9	1.2	700	22	AAU38831	Human polypeptide
29	9	1.2	717	22	AAU40617	Human polypeptide
30	9	1.2	861	23	ABP60988	Novel human protei
31	9	1.2	922	23	ABP60990	Novel human protei
32	9	1.2	1274	22	AAU47329	FCR6. Homo sapie
33	8	1.1	55	22	AAO10097	Human polypeptide
34	8	1.1	63	21	AA51734	Human secreted pro
35	8	1.1	67	21	AAU48327	Arabidopsis thalia
36	8	1.1	67	21	AAU86322	Human secreted pro
37	8	1.1	67	21	AAU86488	Human gene 58-enco
38	8	1.1	68	21	AAU11905	Arabidopsis thalia
39	8	1.1	71	21	AAU48326	Arabidopsis thalia
40	8	1.1	72	21	AAU11904	Arabidopsis thalia
41	8	1.1	74	21	AAU00565	Human secreted pro
42	8	1.1	78	21	AAU48325	Arabidopsis thalia
43	8	1.1	80	21	AAU11903	Arabidopsis thalia
44	8	1.1	86	22	AAU17926	Novel human respir
45	8	1.1	88	22	AAU89233	Human secreted pro

ALIGNMENTS

RESULT 1

AAU70539
ID AAU70539 standard; Protein: 569 AA.

XX AAU70539;

XX 04-JUL-2000 (first entry)

XX Human Factor 8 Homologue.

XX Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
XX cerebroprotective; therapeutic; coagulation related disorder;
XX haemophilia; stroke; screening.

XX Homo sapiens.

XX WO200012532-A1.

XX 09-MAR-2000.

XX 20-AUG-1999; 99WO-US19047.

XX 31-AUG-1998; 98US-0098521.

XX (ELIL) LILLY & CO ELI.

XX Rostock PRJ, Su W, Li XM;

XX WPI; 2000-256580/22.

XX N-PSDB; AAZ51872.

XX Factor 8 homolog polypeptides and nucleic acids encoding them for
XX treating coagulation related disorders such as hemophilia and stroke

XX PS Claim 3; Page 64-66; 68pp; English.

XX CC The present sequence is a human Factor 8 homologue (F8H),

CC a coagulation cofactor which is selectively expressed in

CC haematopoietic, heart and reproductive tissues. It has haemostatic and

CC cerebroprotective activities. The F8H contains a Factor 5/8 signature

CC and is useful as a therapeutic for treating coagulation related diseases

CC such as haemophilia and stroke. The nucleic acid is useful as

CC hybridisation probe and amplification primer for detecting deficiencies

CC in the level of F8H mRNA, for screening F8H gene mutations and for

CC monitoring regulation of gene expression. Fragments of the nucleic acid

CC are also useful as diagnostic probes and primers, and can be used in

CC screening methods such as those using DNA chips. The present sequence is

CC also useful as a target to screen therapeutically useful modulators

CC of the F8H.

XX SQ Sequence 669 AA;

Query Match 91.8%; Score 669; DB 21; Length 669;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 MGERVIRKFGDFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGGLQNMHSTESKNEIT 120

Db 1 MGERVIRKFGDFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGGLQNMHSTESKNEIT 60

QY 121 LLFMSGIHVSGRGFLASVVDKQDLITCLDTASNFLEPESKYPACGLLPFAEISGTI 180

Db 61 LLFMSGIHVSGRGFLASVVDKQDLITCLDTASNFLEPESKYPACGLLPFAEISGTI 120

QY 181 PHGYRDSPLCMAGHAGVSVNTLGGQISVWISKGIPIYESSLANNVTSVVGHLSTSLFT 240

Db 121 PHGYRDSPLCMAGHAGVSVNTLGGQISVWISKGIPIYESSLANNVTSVVGHLSTSLFT 180

QY 241 FRTSCYGTGLMGESGVADPQITASSVLEWTDHTCQNSWPKPKARLKKPGPPWAAFAFD 300

Db 181 FRTSCYGTGLMGESGVADPQITASSVLEWTDHTCQNSWPKPKARLKKPGPPWAAFAFD 240

QY 301 EYQWLQIDLNEKKITGTTGSTMVEHNYYSVAYRIYSDGOKWTVYRPGVEQDKIF 360

Db 241 EYQWLQIDLNEKKITGTTGSTMVEHNYYSVAYRIYSDGOKWTVYRPGVEQDKIF 300

QY 361 QGNKYDHODVRNFPPIIARFIRVNTQWQKIAMKMLLGCQFIPKGRPPKLTQPPPP 420

Db 301 QGNKYDHODVRNFPPIIARFIRVNTQWQKIAMKMLLGCQFIPKGRPPKLTQPPPP 360

QY 421 RNSNDLKNTAPPKTAGRAPKFTQPLQPRSSNEPPAQTETASPDIRNTVTPNVTKD 480

Db 361 RNSNDLKNTAPPKTAGRAPKFTQPLQPRSSNEPPAQTETASPDIRNTVTPNVTKD 420

QY 481 VALAAVLPVLMVLTLLILVCAWHNRNKKTEGTYDLPYDRAGWKGKQFLPAK 540

Db 421 VALAAVLPVLMVLTLLILVCAWHNRNKKTEGTYDLPYDRAGWKGKQFLPAK 480

QY 541 AVDHEETPVYSSSEVNHLSPREVTTVLQADSAEYAQPLVGGIVGTLHORSTFKPEEGKE 600

Db 481 AVDHEETPVYSSSEVNHLSPREVTTVLQADSAEYAQPLVGGIVGTLHORSTFKPEEGKE 540

QY 601 AGYADLPYNPQGVYHAYAEPLDITGPEYATPIIMDSMGHPTTSVGQPSSTSTFKATGN 660

Db 541 AGYADLPYNPQGVYHAYAEPLDITGPEYATPIIMDSMGHPTTSVGQPSSTSTFKATGN 600

QY 661 OPPPLVGYNTLLSNTDSCSAQAQYDTPKAGKPGCLPADVLVYQVQSTQEVSGAGRDG 720

Db 601 OPPPLVGYNTLLSNTDSCSAQAQYDTPKAGKPGCLPADVLVYQVQSTQEVSGAGRDG 660

QY 721 ECDVFEKIL 729

Db 661 ECDVFEKIL 669

RESULT 2

AAG75450

XX ID AAG75450 standard; Protein; 583 AA.

XX AC AAG75450;

XX DT 03-SEP-2001 (first entry)

XX DE Human colon cancer antigen protein SEQ ID NO:6214.

XX KW Human colon cancer; colon cancer antigen; diagnosis; detection;

XX KW colorectal carcinoma.

XX OS Homo sapiens.

XX PN WO200122920-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX PR 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX DR WPI; 2001-235357/24.

XX DR N-PSDB; AAH34855.

XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

XX PS useful for preventing, diagnosing and/or treating colorectal cancers -

Claim 11; Page 7657-7660; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers, AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention.

N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.

XX SQ Sequence 583 AA;

Query Match 44.7%; Score 326; DB 22; Length 583;

Best Local Similarity 99.6%; Pred. No. 4.9e-307;

Matches 526; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPLFLLLLVLLLLLEDAGAGCGGHTVLGPGESGTLTSINYPOTYNSTVCEWEIRVK 60

Db 25 MPLFLLLLVLLLLLEDAGAGCGGHTVLGPGESGTLTSINYPOTYNSTVCEWEIRVK 84

QY 61 MGERVIRKFGDFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGGLQNMHSTESKNEIT 120

Db 85 MGERVIRKFGDFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGGLQNMHSTESKNEIT 144

QY 121 LLFMSGIHVSGRGFLASVVDKQDLITCLDTASNFLEPESKYPACGLLPFAEISGTI 180

Db 145 LLFMSGIHVSGRGFLASVVDKQDLITCLDTASNFLEPESKYPACGLLPFAEISGTI 204

QY 181 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISGIPYESSLANNVTSVVGHLSTSLFT 240
 DB 205 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISGIPYESSLANNVTSVVGHLSTSLFT 264
 QY 241 FKTSGCYGTGLMGESGVADPOITASSVLEWTDHTGOENSWKPKKARLKKPGPPWAAAFATD 300
 DB 265 FKTSGCYGTGLMGESGVADPOITASSVLEWTDHTGOENSWKPKKARLKKPGPPWAAAFATD 324
 QY 301 EYQWLQDLNKEKKITGIITGSTWVEHNYVYSAIRILYSDGQKWTYRPGVEQDKIF 360
 DB 325 EYQWLQDLNKEKKITGIITGSTWVEHNYVYSAIRILYSDGQKWTYRPGVEQDKIF 384
 QY 361 QGNKDYHODVNNELPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPPKLTQPPPP 420
 DB 385 QGNKDYHODVNNELPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPPKLTQPPPP 444
 QY 421 RNSNDLKNWTPAPPIAKGRAPKFTQLOPRSSNEFFPAQTEQTTASPDTRNTVTPNVTKD 480
 DB 445 RNSNDLKNWTPAPPIAKGRAPKFTQLOPRSSNEFFPAQTEQTTASPDTRNTVTPNVTKD 504
 QY 481 VALAAVLVPVLVWVLTLLILVCAWHNRNKKTEGYDLPYWDRA 528
 DB 505 VALAAVLVPVLVWVLTLLILVCAWHNRNKKTEGYDLPYWDRA 552

RESULT 3

AAE22716

ID AAE22716 standard; Protein: 385 AA.

XX AC AAE22716;

DT 09-AUG-2002 (first entry)

DE Human neuropilin-Hy2 protein.

XX

KW Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;

KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;

KW thrombocytopenia; memory; platelet; plastic anaemia; antiinflammatory;

KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;

KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;

KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;

KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;

KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;

KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;

KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;

KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;

KW neurotropic; neuroprotective; vulnary; anticonvulsant; antiparasitic;

KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;

KW immunosuppressive; chromosome 6q21.

XX OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide 1..20

FT Protein /label= Signal_peptide

FT 61..385

FT /note= "Human mature neuropilin-Hy2 protein"

XX

PN WO200222815-A1.

XX

PD 21-MAR-2002.

XX

PF 12-SEP-2001; 2001WO-US28488.

XX

PR 11-SEP-2000; 2000US-0659671.

XX

PR 06-SEP-2001; 2001US-317902P.

XX

PA (HYSE-) HYSEQ INC.

XX

XX Tang YT;

XX

DR WPI: 2002-393966/42.

XX

DR N-PSDB; AAD35994.

XX

XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects
 XX Claim 3; Page 128-130; 152pp; English.
 XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy2
 CC protein. Neuropilin-Hy2 gene is located on chromosome 6q21.

XX Sequence 385 AA;

Query Match 35.0%; Score 255; DB 23; Length 385;
 Best Local Similarity 100.0%; Pred. No. 2.4e-238;
 Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPLFLLLLVLLLEDDAGAQOQCGHTVLGPESGTLTSINYPQTPNSTVCEWEIRVK 60

DB 1 MPLFLLLLVLLLEDDAGAQOQCGHTVLGPESGTLTSINYPQTPNSTVCEWEIRVK 60

QY 61 MGERVRIKFGDFDIEDSDSCHENYLRVINGIGVSRTEIGKYCGLGLQNMHSTESKGNIT 120

DB 61 MGERVRIKFGDFDIEDSDSCHENYLRVINGIGVSRTEIGKYCGLGLQNMHSTESKGNIT 120

QY 121 LFMSSGIVHSGRGFLASYSVIDKDLITCLDTPASNFLEPFESKYPAGCLLPFAETSGTI 180

DB 121 LFMSSGIVHSGRGFLASYSVIDKDLITCLDTPASNFLEPFESKYPAGCLLPFAETSGTI 180

QY 181 PHGYRDSPLCMAGVHAGVSVNTLGGQISVWISKGIPIYESSLANNVTSVVGHLSTSLFT 240

DB 181 PHGYRDSPLCMAGVHAGVSVNTLGGQISVWISKGIPIYESSLANNVTSVVGHLSTSLFT 240

QY 241 FKTSGCYGTGLMGESG 255

DB 241 FKTSGCYGTGLMGESG 255

RESULT 4

AAU79460

ID AAU79460 standard; Protein: 385 AA.

XX AC AAU79460;

XX

XX 15-JUL-2002 (first entry)

XX Human Neuropilin-Hy2.

XX

XX Human; neuropilin-Hy2; chromosome 6q21; neuronal growth;

XX KW

nerve regeneration; neurodegenerative disease; learning disorder;
memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
organ growth; nervous system lesion; cancer; cell proliferation;
cell differentiation; stem cell growth factor activity;
Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
reperfusion; food supplement; DNA microarray.

OS Homo sapiens.

XX Key Location/Qualifiers
FH Peptide 1..20
FT /label= signal_peptide
FT Protein 21..385
FT /label= Mature_neuropilin_Hy2

XX WO200222780-A2.

XX 21-MAR-2002.

XX 11-SEP-2001; 2001WO-US28590.

XX 11-SEP-2000; 2000US-0659671.

PR 06-SEP-2001; 2001US-0659671.

XX (TANG/) TANG T Y.

XX Tang TY;

XX WPI; 2002-351881/38.

DR N-PSDB; ABK49567.

XX New neuropilin-like polypeptides for diagnosing, preventing and
PT treating neurological conditions and disorders, cancers, and for
PT inducing angiogenesis and neovascularisation.

XX Claim 3; Page 123-125; 144pp; English.

XX The invention relates to an isolated neuropilin-like polypeptide
CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
CC encoding the proteins and the coding regions of the cDNAs. Also included
CC is a nucleic acid array comprising the cDNAs attached to a surface used
CC for detecting full-matches or mismatches to the cDNAs. The genes
CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
CC The nucleic acid array is useful for detecting full-matches or mismatches
CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
CC in modulating neuronal growth, regenerative capacity, treating
CC neurodegenerative diseases, learning and memory disorders, diagnosing and
CC mapping genetic neuronal defects and degenerative diseases like
CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
CC and organ growth and development (e.g. the heart). The nervous system
CC disorders include lesions of central or peripheral nervous systems,
CC including traumatic lesions, ischaemic lesions, infectious lesions,
CC degenerative lesions, lesions associated with nutritional diseases or
CC disorders, neurological lesions, and lesions caused by toxic substances.
CC The neuropilin-like proteins and cDNAs are also useful as markers for
CC cancers. The neuropilin-like proteins are useful for regulating cell
CC proliferation, cell differentiation, stem cell growth factor activity,
CC for inducing proliferation of neural cells, regeneration of nerve and
CC brain tissue, for treatment of central and peripheral nervous system
CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopaenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues and as a food supplement or molecular
CC weight marker. The cDNAs are useful in gene identification, genome
CC mapping, transgenics, as hybridisation probes, for primer design, for
CC gene chips and as a DNA antigen. The present sequence represents
CC neuropilin-Hy2.

XX Sequence 385 AA;

Query Match 35.0%; Score 255; DB 23; Length 385;
Best Local Similarity 100.0%; Pred. No. 2.4e-238;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPFLLLLVLLLEDAGAGGCGHTVLGPESGTLTSINYPOTPNSTVCEWEIRVK 60
DB 1 MPFLLLLVLLLEDAGAGGCGHTVLGPESGTLTSINYPOTPNSTVCEWEIRVK 60

QY 61 MGERVRIKFGDFIEDSDSCHFNRLIYNGIGVSRTEIGKYCGLGLQNNHSIESKGNFT 120
DB 61 MGERVRIKFGDFIEDSDSCHFNRLIYNGIGVSRTEIGKYCGLGLQNNHSIESKGNFT 120

QY 121 LLFMSGHVSGRGFLASYSVIDKQDLITCLDTASNFLEPFESKYCPAGCLLPFAISGTI 180
DB 121 LLFMSGHVSGRGFLASYSVIDKQDLITCLDTASNFLEPFESKYCPAGCLLPFAISGTI 180

QY 181 PHGYRDSPLCMAGVHAGVWSNTLGGQISVVISKGIPIYESSLANNVTSVGHLSLTSLFT 240
DB 181 PHGYRDSPLCMAGVHAGVWSNTLGGQISVVISKGIPIYESSLANNVTSVGHLSLTSLFT 240

QY 241 FKTSGCYGTLCWESG 255

DB 241 FKTSGCYGTLCWESG 255

RESULT 5

ABB97386

ID ABB97386 standard; Protein; 385 AA.

XX ABB97386;

DT 27-JUN-2002 (first entry)

DE Novel human protein SEQ ID NO: 654.

Human; antianaemic; vulnary; antiinflammatory; immunomodulator;
antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
neuroprotective; antiparkinsonian; protein therapy; EST;
expressed sequence tag.

OS Homo sapiens.

XX WO200222660-A2.

XX 21-MAR-2002.

XX 10-SEP-2001; 2001WO-US26015.

XX 11-SEP-2000; 2000US-0659671.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

DR WPI: 2002-292408/33.

DR N-PSDB; ABN32572.

XX An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis -

XX Claim 20; SEQ ID NO 654; 509pp; English.

XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate haemostasis or inhibit e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.

CC Parkinson's disease. The present sequence is a protein of the invention.

XX
SQ Sequence 385 AA;
Query Match 35.0%; Score 255; DB 23; Length 385;
Best Local Similarity 100.0%; Pred. No. 2.4e-238;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPFLFLLLVLLLEDAAGQGGCGHTVLGPESGTLTSINYPOTYPNSTVCEWEIRVK 60
DB 1 MPFLFLLLVLLLEDAAGQGGCGHTVLGPESGTLTSINYPOTYPNSTVCEWEIRVK 60
QY 61 MGERVRIKFGDFIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGLGLQNMHSIESKGNIT 120
DB 61 MGERVRIKFGDFIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGLGLQNMHSIESKGNIT 120
QY 121 LFWMSGIHVSGRGLFASVVDKQDLITCLDTASNFLPEPESKYCPAGCLLPFAISGTI 180
DB 121 LFWMSGIHVSGRGLFASVVDKQDLITCLDTASNFLPEPESKYCPAGCLLPFAISGTI 180
QY 181 PHGYRDSPLCMAGVHAGVSNLTGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFT 240
DB 181 PHGYRDSPLCMAGVHAGVSNLTGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFT 240
QY 241 FKTSCYGTGLMESG 255
DB 241 FKTSCYGTGLMESG 255

RESULT 6

AAE22721
ID AAE22721 standard; Protein; 365 AA.

XX AAE22721;
AC AAE22721;

XX 09-AUG-2002 (first entry)
XX Human neuropilin-Hy2 mature protein.

XX Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neutropenic; neuroprotective; vulnary; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive.

XX Homo sapiens.

XX WO200222815-A1.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28488.

XX 11-SEP-2000; 2000US-0659671.

XX 06-SEP-2001; 2001US-317902P.

XX (HYSE-) HYSEQ INC.

XX Tang YT;

XX WPI; 2002-393966/42.

XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
XX useful for treating neurodegenerative diseases e.g. Alzheimer's

PT disease, and for diagnosing and mapping genetic neuronal defects
XX Disclosure; Page 131-132; 152pp; English.

XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.
CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat
CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
CC nocturnal haemoglobinuria and is used in nerve tissue growth or
CC regeneration, in wound healing, tissue repair and replacement and in
CC healing of bones, incisions and ulcers. Compositions comprising the
CC sequences of the invention are useful for treating diseases of peripheral
CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
CC stroke, ulcers, immune deficiencies and immune disorders. Infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy2
XX mature protein.

SQ Sequence 365 AA;

Query Match 32.2%; Score 235; DB 23; Length 365;

Best Local Similarity 100.0%; Pred. No. 5.6e-219;

Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 OQDGGCHVTLPESGTLTSINYPOTYPNSTVCEWEIRVKMGRIKFGDFIEDSDSC 80

DB 1 OQDGGCHVTLPESGTLTSINYPOTYPNSTVCEWEIRVKMGRIKFGDFIEDSDSC 60

QY 81 HFNLYRIYNGIGVSRTEIGKYCGLGLQNMHSIESKGNITLLFMSGIHYSGRGLASYSV 140

DB 61 HFNLYRIYNGIGVSRTEIGKYCGLGLQNMHSIESKGNITLLFMSGIHYSGRGLASYSV 120

QY 141 IDKQDLITCLDTASNFLPEPESKYCPAGCLLPFAISGTIPHYGRDSSPLCMAGVHAGV 200

DB 121 IDKQDLITCLDTASNFLPEPESKYCPAGCLLPFAISGTIPHYGRDSSPLCMAGVHAGV 180

QY 201 SNTLGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFTFTKTCYGTGLMESG 255

DB 181 SNTLGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFTFTKTCYGTGLMESG 235

RESULT 7

ABG01298

ID ABG01298 standard; Protein; 197 AA.

XX ABG01298;

XX 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #1289.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS65485.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX Claim 20; SEQ ID NO 31657; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG0377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 197 AA;
Query Match 27.0%; Score 197; DB 22; Length 197;
Best Local Similarity 100.0%; Pred. No. 2.3e-182; Mismatches 0; Indels 0; Gaps 0;
Matches 197; Conservative 0;
QY 533 MKQFLPAKAVDHEETPVYSSSEVNHLSPREVTTLQADSAYEAPLVGGIVGTLHQKST 592
DB 1 MKQFLPAKAVDHEETPVYSSSEVNHLSPREVTTLQADSAYEAPLVGGIVGTLHQKST 60
QY 593 FKPEEGKAGYADLPYNSPGQEVHAYAEPLPTGPEYATPIIMDSGHPTTSVQGPST 652
DB 61 FKPEEGKAGYADLPYNSPGQEVHAYAEPLPTGPEYATPIIMDSGHPTTSVQGPST 120
QY 653 STFKATGNPPPLVGYNTLLSRTDSCSSAQAYQDTPKAGKPLGAPDELVYQVPQSTOE 712
DB 121 STFKATGNPPPLVGYNTLLSRTDSCSSAQAYQDTPKAGKPLGAPDELVYQVPQSTOE 180
QY 713 VSGAGRDGECDFVKEIL 729
DB 181 VSGAGRDGECDFVKEIL 197
RESULT 8
ID ABB42217
XX ABB42217 standard; Peptide; 75 AA.
XX ABB42217;
XX 04-FEB-2002 (first entry)
XX

DE Peptide #9723 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
XX WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00669.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 27; SEQ ID NO 34852; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 75 AA;
Query Match 10.3%; Score 75; DB 22; Length 75;
Best Local Similarity 100.0%; Pred. No. 2e-64;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 24 DCGHTVLGPESGTLTSINYPOTYPNSTVCWEIRVKMGRIKFGDFDIEDSDSCHFN 83
DB 1 DCGHTVLGPESGTLTSINYPOTYPNSTVCWEIRVKMGRIKFGDFDIEDSDSCHFN 60
QY 84 YLRIYNGIGVSRTEI 98
DB 61 YLRIYNGIGVSRTEI 75
RESULT 9
ID AAM63102
XX AAM63102 standard; Protein; 75 AA.
XX AC
XX AAM63102;
XX 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 35207.
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer.
XX Homo sapiens.
XX WO200157275-A2.
XX

```
XX PD. 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX PS Example 4; SEQ ID NO: 35207; 650pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of
XX CC the probes of the invention.
XX SQ Sequence 75 AA;
XX Query Match 10.3%; Score 75; DB 22; Length 75;
XX Best Local Similarity 100.0%; Pred. No. 2e-64;
XX Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 24 DGCCHTVLGPESGTLTINYPQTPNVCWEIRVKGKGVRIKFGDFDIEDSDSCHFN 83
Db 1 DGCCHTVLGPESGTLTINYPQTPNVCWEIRVKGKGVRIKFGDFDIEDSDSCHFN 60
QY 84 YLRIYNGIGVSRTEI 98
Db 61 YLRIYNGIGVSRTEI 75
RESULT 10
AAM75913
ID AAM75913 standard; Protein; 75 AA.
XX AC AAM75913;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 36219.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
```

```
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 36219; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention.
XX SQ Sequence 75 AA;
XX Query Match 10.3%; Score 75; DB 22; Length 75;
XX Best Local Similarity 100.0%; Pred. No. 2e-64;
XX Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 24 DGCCHTVLGPESGTLTINYPQTPNVCWEIRVKGKGVRIKFGDFDIEDSDSCHFN 83
Db 1 DGCCHTVLGPESGTLTINYPQTPNVCWEIRVKGKGVRIKFGDFDIEDSDSCHFN 60
QY 84 YLRIYNGIGVSRTEI 98
Db 61 YLRIYNGIGVSRTEI 75
RESULT 11
AAM36024
ID AAM36024 standard; Protein; 75 AA.
XX AC AAM36024;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #10061 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
```

XX Claim 27; SEQ ID No 36293; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP: see AAI1315-AAI57546). The present sequence is a peptide encoded by one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders.

XX Sequence 75 AA;

SQ Query Match 10.3%; Score 75; DB 22; Length 75;
Best Local Similarity 100.0%; Pred. No. 2e-64;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 DCGCHTVLGPESGTLTSINYPQTPNSVCEWEIRKMGVRIRKFGDFDIEDSDSCHFN 83
Db 1 DCGCHTVLGPESGTLTSINYPQTPNSVCEWEIRKMGVRIRKFGDFDIEDSDSCHFN 60

QY 84 YLRIYNGIGVSRTEI 98
Db 61 YLRIYNGIGVSRTEI 75

RESULT 12
ABG45349
ID ABG45349 standard; Peptide: 75 AA.
XX AC ABG45349;
XX DT 19-AUG-2002 (first entry)
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 35014.
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberculous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX OS Homo sapiens.
XX WO200186003-A2.
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US00665.
XX PR 04-FEB-2000; 2000US-180312P.
XX PR 26-MAY-2000; 2000US-207456P.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-234687P.
XX PR 27-SEP-2000; 2000US-236359P.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples -
XX Claim 27; SEQ ID No 35014; 634pp; English.
XX The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarray having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberculous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 75 AA;

QY Query Match 10.3%; Score 75; DB 23; Length 75;
Best Local Similarity 100.0%; Pred. No. 2e-64;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 DCGCHTVLGPESGTLTSINYPQTPNSVCEWEIRKMGVRIRKFGDFDIEDSDSCHFN 83
Db 1 DCGCHTVLGPESGTLTSINYPQTPNSVCEWEIRKMGVRIRKFGDFDIEDSDSCHFN 60

QY 84 YLRIYNGIGVSRTEI 98
Db 61 YLRIYNGIGVSRTEI 75

RESULT 13
AAE22720
ID AAE22720 standard; peptide: 20 AA.
XX AC AAE22720;
XX DT 09-AUG-2002 (first entry)
XX DE Human neuropilin-Hy2 signal peptide.
XX Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation; neurodegenerative disease; Alzheimer's disease; learning; angioneitis; thrombocytopaenia; memory; platelet; placental anaemia; anti-inflammatory; ophthalmological; paroxysmal nocturnal haemoglobinuria; procoagulant; wound healing; tissue repair; Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord; cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis; human immunodeficiency virus; HIV; autoimmune disorder; dermatological;

KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neotrophic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive.
XX Homo sapiens.
XX OS
XX PN WO200222815-A1.
XX PD
XX PF 21-MAR-2002.
XX PP 12-SEP-2001; 2001WO-US28488.
XX PR 11-SEP-2000; 2000US-0659671.
XX PR 06-SEP-2001; 2001US-317902P.
XX PA (HYSE-) HYSEQ INC.
XX PT Tang YT;
XX PI WPI; 2002-393966/42.
XX DR Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
XX PT useful for treating neurodegenerative diseases e.g. Alzheimer's
XX PT disease, and for diagnosing and mapping genetic neuronal defects
XX PS Disclosure; Page 130-131; 152pp; English.
XX CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
XX CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
XX CC like polypeptides and polynucleotides are useful in modulating neuronal
XX CC growth regenerative capacity, treating neurodegenerative diseases,
XX CC diagnosing and mapping genetic neuronal defects and degenerative diseases
XX CC like Alzheimer's disease and for treating learning and memory disorders.
XX CC They are also useful for inducing angiogenesis, neovascularisation, as
XX CC well as organ growth and development e.g. heart and other tissues.
XX CC Antagonists of neuropilin-like polypeptides are useful for treating
XX CC cancers and other malignant diseases. Neuropilin is used to treat
XX CC platelet disorders e.g. thrombocytopenia, plastic anaemia and paroxysmal
XX CC nocturnal haemoglobinuria and is used in nerve tissue growth or
XX CC regeneration, in wound healing, tissue repair and replacement and in
XX CC healing of bones, incisions and ulcers. Compositions comprising the
XX CC sequences of the invention are useful for treating diseases of peripheral
XX CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
XX CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
XX CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
XX CC stroke, ulcers, immune deficiencies and immune disorders, infections by
XX CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
XX CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
XX CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
XX CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
XX CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
XX CC inflammatory eye disease. The nucleic acids of the invention are used in
XX CC gene therapy techniques. The present sequence is human neuropilin-Hy2
XX CC signal peptide.
XX SQ Sequence 20 AA;
Query Match 2.7%; Score 20; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.3e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 MFLFLLLVLLVLLLEDAGA 20
Db 1 MFLFLLLVLLVLLLEDAGA 20
RESULT 14
AAU74424
ID AAU74424 standard; Protein; 629 AA.
XX

AC AAU74424;
XX 23-APR-2002 (first entry)
XX Mouse protein sequence #3, related to isolation of genes within SLE-1B.
DE DE
XX Mouse; systemic lupus erythematosus 1B; SLE-1B; dermatological;
KW antiinflammatory; immunosuppressive; systemic autoimmune disorder;
KW signalling lymphocyte activation molecule; SLAM; lymphocyte antigen 9;
KW Ly-9; 2B4; natural killer cell receptor; CD48; CD84; LY108; CS1; DEDD;
KW NIT1; upstream transcription factor 1; USF 1; GOLGA4; immune tolerance.
XX OS Mus musculus.
XX PN WO20018200-A2.
XX PD 22-NOV-2001.
XX PF 17-MAY-2001; 2001WO-US16051.
XX PR 17-MAY-2000; 2000US-204963P.
XX PR 21-SEP-2000; 2000US-234437P.
XX PA (TEXA) UNIV TEXAS SYSTEM.
XX PI Wakeland EK, Wandstrat A, Morel L;
XX PT WPI; 2002-066695/09.
XX DR N-PSDB; AAS20419.
XX PT Screening for susceptibility to systemic autoimmune disorder by
XX PT screening for a mutation within the systemic lupus erythematosus-1B
XX PT loci
XX PS Disclosure; Page 115-117; 128pp; English.
XX CC The present invention relates to a new method for screening for
XX CC susceptibility to a systemic autoimmune disorder. The method comprises
XX CC screening for at least one mutation within the systemic lupus
XX CC erythematosus (SLE)-1B loci. Screening for susceptibility to autoimmune
XX CC disorders such as systemic lupus erythematosus involves screening for at
XX CC least one mutation in a gene or genes with the SLE-1B loci such as a gene
XX CC encoding signalling lymphocyte activation molecule (SLAM), lymphocyte
XX CC antigen (Ly)-9, 2B4 (a natural killer cell receptor), CD48, CD84, LY108,
XX CC CS1, DEDD, NIT1, upstream transcription factor (USF)1, GOLGA4. The method
XX CC of the invention is useful for treating SLE and involves administering
XX CC a construct comprising a wild-type sequence encoding any one of the
XX CC above mentioned genes. Gene therapy also involves the use of antisense
XX CC constructs or ribozymes directed against the above mentioned genes for
XX CC treating SLE.
XX CC Note: The present protein sequence represents a mouse protein of the
XX CC invention that relates to the isolation of genes within SLE-1B that
XX CC mediate a break in immune tolerance. This sequence is shown in the
XX CC sequence listing but is not further defined in the specification.
XX SQ Sequence 629 AA;
Query Match 1.4%; Score 10; DB 23; Length 629;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 5 LLLLVLLVLLLL 14
Db 435 LLLLVLLVLLLL 444
RESULT 15
AAW70900
ID AAW70900 standard; Protein; 677 AA.
XX AC AAW70900;
XX 14-OCT-1998 (first entry)
XX

```
XX DE Human beta galactosidase protein.
XX DE
XX KW Retroviral vector; gene delivery vehicle; expression;
KW non-immunogenic selectable marker; gene therapy; activation;
KW human; beta galactosidase.
XX OS Homo sapiens.
XX PN WO9830709-A2.
XX PD 16-JUL-1998.
XX PF 14-JAN-1998; 98WO-US00715.
XX PR 13-JAN-1998; 98US-0038339.
XX PR 14-JAN-1997; 97US-0035473.
XX PR 27-FEB-1997; 97US-0038339.
XX PA (CHAD/) CHADA S.
XX PA (JOLLY) JOLLY D J.
XX PA (MOOR/) MOORE M D.
XX PI Chada S, Jolly DJ, Moore MD;
XX WPI: 1998-399153/34.
XX DR N-PSDB; AAV42728.
XX PT Non-immunogenic pro:drug activating enzyme(s) and selectable
PT marker(s) - are used in gene therapy for the treatment of a wide
PT variety of disorder(s)
XX Example 2; Fig 5A-B; 121pp; English.
XX The present sequence represents human beta galactosidase and is used as
CC a marker in the retroviral vector of the invention. The specification
CC describes a gene delivery vehicle which directs expression of a
CC non-immunogenic selectable marker or molecule which is capable of
CC activating a previously inactive compound. Vectors expressing the
CC markers and a heterologous sequence are useful in gene therapy. The
CC vectors can be used to deliver a molecule into a target area where it
CC may cause the activation of a previously inactive substance.
XX SQ Sequence 677 AA;
Query Match 1.48; Score 10; DB 19; Length 677;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 LLLLVLLLL 14
| | | | |
Db 9 LLLLVLLLL 18
Search completed: January 21, 2003, 09:57:14
Job time : 30.4848 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:53:55 ; Search time 25.5152 Seconds
(without alignments)
3410.229 Million cell updates/sec

Title: US-10-060-830-1114

Perfect score: 653

Sequence: 1 MPLFLLLVLLVLLLEDAGA.....TQEVSGAGRGDCDVFKEIL 653

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID\$2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	631	96.6	669	21	Human Factor 8 Hom
2	228	34.9	593	22	Human colon cancer
3	197	30.2	197	22	Novel human diago
4	157	24.0	365	23	Human neuropilin-H
5	157	24.0	385	23	Human neuropilin-H
6	157	24.0	385	23	Human neuropilin-H
7	157	24.0	385	23	Human neuropilin-H
8	20	3.1	20	23	Human neuropilin-H
9	10	1.5	629	23	Mouse protein sequ
10	10	1.5	677	19	Human beta galacto

11	10	1.5	696	21	Human prostate can
12	9	1.4	73	22	Propionibacterium
13	9	1.4	89	23	Human albumin fusi
14	9	1.4	89	23	Human gene 1 encod
15	9	1.4	150	23	Novel secreted pro
16	9	1.4	150	23	Human secreted cyt
17	9	1.4	209	23	Novel secreted pro
18	9	1.4	276	19	H. pylori GHPO 174
19	9	1.4	377	22	Human polypeptide
20	9	1.4	377	23	Human WKU-1 polype
21	9	1.4	441	22	Human polypeptide
22	9	1.4	700	22	Human EST encoded
23	9	1.4	700	22	Human polypeptide
24	9	1.4	717	22	Human polypeptide
25	9	1.4	861	23	Novel human protei
26	9	1.4	922	23	Novel human protei
27	9	1.4	1274	22	FCTR6. Homo sapie
28	8	1.2	55	22	Human polypeptide
29	8	1.2	63	21	Human secreted pro
30	8	1.2	67	21	Arabidopsis thalia
31	8	1.2	67	21	Human secreted pro
32	8	1.2	67	21	Human gene 58-enco
33	8	1.2	68	21	Arabidopsis thalia
34	8	1.2	71	21	Arabidopsis thalia
35	8	1.2	72	21	Arabidopsis thalia
36	8	1.2	74	21	Human secreted pro
37	8	1.2	78	21	Arabidopsis thalia
38	8	1.2	80	21	Arabidopsis thalia
39	8	1.2	86	22	Novel human respir
40	8	1.2	88	22	Human secreted pro
41	8	1.2	95	21	Arabidopsis thalia
42	8	1.2	102	21	Plant SDF encoded
43	8	1.2	103	21	Arabidopsis thalia
44	8	1.2	103	21	C. elegans insulin
45	8	1.2	105	12	B.subtilis rib ORF

ALIGNMENTS

RESULT 1

AAAY70539 standard; Protein; 669 AA.

AC AAAY70539;

XX 04-JUL-2000 (first entry)

XX Human Factor 8 Homologue.

XX Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;

XX cerebroprotective; therapeutic; coagulation related disorder;

XX haemophilia; stroke; screening.

OS Homo sapiens.

XX WO200012532-A1.

XX 09-MAR-2000.

XX 20-AUG-1999; 99WO-US19047.

XX 31-AUG-1998; 98US-0098521.

XX (ELIL) LILLY & CO ELI.

XX Rostock PRJ, Su W, Li XM;

XX WPI; 2000-256580/22.

XX N-PSDB; AA251872.

XX Factor 8 homolog polypeptides and nucleic acids encoding them for

XX treating coagulation related disorders such as hemophilia and stroke

XX Claim 3; Page 64-66; 68pp; English.

XX The present sequence is a human Factor 8 homologue (F8H),

XX a coagulation cofactor which is selectively expressed in

CC haematopoietic, heart and reproductive tissues. It has haemostatic and

CC cerebroprotective activities. The F8H contains a Factor 5/8 signature

CC and is useful as a therapeutic for treating coagulation related diseases

CC such as haemophilia and stroke. The nucleic acid is useful as

CC hybridisation probe and amplification primer for detecting deficiencies

CC in the level of F8H mRNA, for screening F8H gene mutations and for

CC monitoring regulation of gene expression. Fragments of the nucleic acid

CC are also useful as diagnostic probes and primers, and can be used in

CC screening methods such as those using DNA chips. The present sequence is

CC also useful as a target to screen therapeutically useful modulators

CC of the F8H.

XX Sequence 669 AA;

SQ Query Match 96.6%; Score 631; DB 21; Length 669;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 631; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GKYCGLGLQMNHSIESKGNIEITLLFMSGIHVSGRGLASYSVIDKQDLITCLDTASNFLE 82

DB 39 GKYCGLGLQMNHSIESKGNIEITLLFMSGIHVSGRGLASYSVIDKQDLITCLDTASNFLE 98

QY 83 PFESKYPAGCGLLPFAEISGTIPHGVRDSSPLCMAGVHAGVSVNTLGGQISVVISKGIPY 142

DB 99 PFESKYPAGCGLLPFAEISGTIPHGVRDSSPLCMAGVHAGVSVNTLGGQISVVISKGIPY 158

QY 143 YESSLANNTSVVGHLSLTSLFTFKTSGCYGTGLMESGVIADPQITASSVLEWTDHTGQEN 202

DB 159 YESSLANNTSVVGHLSLTSLFTFKTSGCYGTGLMESGVIADPQITASSVLEWTDHTGQEN 218

QY 203 SWKPKARLKKPGPPWAAATDEYQWLQDLNKEKITGIITGTSTWVHNHYVVSAYRIL 262

DB 219 SWKPKARLKKPGPPWAAATDEYQWLQDLNKEKITGIITGTSTWVHNHYVVSAYRIL 278

QY 263 YSDDGQKWTYREPGVEQDKIFQGNKDYHQDVNRNPLPPIIARFIRVNPTQMQKIAMKM 322

DB 279 YSDDGQKWTYREPGVEQDKIFQGNKDYHQDVNRNPLPPIIARFIRVNPTQMQKIAMKM 338

QY 323 ELLGCOFIPKGRPPKLTQPPPRNSNDLKNTTAPPKIAKGRAPKFTQPLQPRSSNEFFPAQ 382

DB 339 ELLGCOFIPKGRPPKLTQPPPRNSNDLKNTTAPPKIAKGRAPKFTQPLQPRSSNEFFPAQ 398

QY 383 TEQTTASPDIRNTTVPNTVKDVALAALVPVLVAVLTLILLYLCANWHNRKKTGTEG 442

DB 399 TEQTTASPDIRNTTVPNTVKDVALAALVPVLVAVLTLILLYLCANWHNRKKTGTEG 458

QY 443 YDLPLYWDAGWNGMKQFUPAKAVDHEETPVRYSSSEVNHLSPREVTTVLQADSAEYAQP 502

DB 459 YDLPLYWDAGWNGMKQFUPAKAVDHEETPVRYSSSEVNHLSPREVTTVLQADSAEYAQP 518

QY 503 LVGGIVGTILHQSTFKPEEGKEAGYADLDYPNSPGQEVYHAYAEPLITGPETATPIIND 562

DB 519 LVGGIVGTILHQSTFKPEEGKEAGYADLDYPNSPGQEVYHAYAEPLITGPETATPIIND 578

QY 563 MSGHPITSVGQSTSTFKATGNQPPPLVGTYNLTLSRTDSCSSAQAQYDTPRAGPGLPA 622

DB 579 MSGHPITSVGQSTSTFKATGNQPPPLVGTYNLTLSRTDSCSSAQAQYDTPRAGPGLPA 638

QY 623 PDELVYQVQSTQSVGAGRDGECDFKEIL 653

DB 639 PDELVYQVQSTQSVGAGRDGECDFKEIL 669

RESULT 2

AAAG75450

ID AAG75450 standard; Protein; 583 AA.

XX

AC AAG75450;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6214.

DE Human: colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma.

XX Homo sapiens.

OS WO200122920-A2.

PN 05-APR-2001.

PD 28-SEP-2000; 2000WO-US26524.

PF 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, Barash SC, Birse CE, Rosen CA;

PI WPI; 2001-235357/24.

XX N-PSDB; AAH34855.

DR Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

XX useful for preventing, diagnosing and/or treating colorectal cancers -

PS Claim 11; Page 7657-7660; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where

CC the proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene

CC therapy and vaccine production. N and P may be used in the prevention,

CC diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders

CC associated with decreased expression by rectifying mutations or deletions

CC in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patients own production of P.

CC Additionally, N may be used to produce the colon cancer-associated Ps,

CC by inserting the nucleic acids into a host cell and culturing the cell

CC to express the proteins. N and P can be used in the prevention, diagnosis

CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

CC and AAG77789 represent sequences used in the exemplification of the

CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for

XX SEQ ID NO:1027 to 1052, 7921 and 7922.

SQ Sequence 583 AA;

Query Match 34.9%; Score 228; DB 22; Length 583;

Best Local Similarity 99.5%; Pred. No. 2.6e-214;

Matches 428; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 23 GKYCGLGLQMNHSIESKGNIEITLLFMSGIHVSGRGLASYSVIDKQDLITCLDTASNFLE 82

DB 123 GKYCGLGLQMNHSIESKGNIEITLLFMSGIHVSGRGLASYSVIDKQDLITCLDTASNFLE 182

QY 83 PFESKYPAGCGLLPFAEISGTIPHGVRDSSPLCMAGVHAGVSVNTLGGQISVVISKGIPY 142

DB 183 PFESKYPAGCGLLPFAEISGTIPHGVRDSSPLCMAGVHAGVSVNTLGGQISVVISKGIPY 242

QY 143 YESSLANNTSVVGHLSLTSLFTFKTSGCYGTGLMESGVIADPQITASSVLEWTDHTGQEN 202

DB 243 YESSLANNTSVVGHLSLTSLFTFKTSGCYGTGLMESGVIADPQITASSVLEWTDHTGQEN 302

QY 203 SWKPKARLKKPGPPWAAATDEYQWLQDLNKEKITGIITGTSTWVHNHYVVSAYRIL 262

DB 303 SWKPKARLKKPGPPWAAATDEYQWLQDLNKEKITGIITGTSTWVHNHYVVSAYRIL 362

QY 263 YSDGQKWTYREPGVEODKIFQGNKDYHQDVNNFLPPIIARFIRVNPQTQOQKIAMKM 322
DB 363 YSDGQKWTYREPGVEODKIFQGNKDYHQDVNNFLPPIIARFIRVNPQTQOQKIAMKM 422
QY 323 ELLGCOFTPKGPPKLTOTPPPPRNSNDLKNTPAPPKIAGRAPKFTQLOPRSSNEFFAQ 382
DB 423 ELLGCOFTPKGPPKLTOTPPPPRNSNDLKNTPAPPKIAGRAPKFTQLOPRSSNEFFAQ 482
QY 383 TEQTTASPDIRNTTTPNVTKDVAALVLPVLMVLTLLILVCAWHNRKXKTEGT 442
DB 483 TEQTTASPDIRNTTTPNVTKDVAALVLPVLMVLTLLILVCAWHNRKXKTEGT 542
QY 443 YDLPYWDRAQ 452
DB 543 YDLPYWDRAQ 552

RESULT 3
ID ABG01298 standard; Protein; 197 AA.
XX
AC ABG01298;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #1289.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS65485.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 31657; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 197 AA;

Query Match 30.2%; Score 197; DB 22; Length 197;
Best Local Similarity 100.0%; Pred. No. 2.3e-184;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 MKQFLPAKAVDHEETPVRYSSSEVNHLSPREVTVLQADSAEYAOQLVGGIVGTLHORST 516
DB 1 MKQFLPAKAVDHEETPVRYSSSEVNHLSPREVTVLQADSAEYAOQLVGGIVGTLHORST 60
QY 517 FKPEEGKEAGYADLDPFYNSPGQEVYHAYAEPLITGPEYATPIIMDSGHPTTSVGQPS 576
DB 61 FKPEEGKEAGYADLDPFYNSPGQEVYHAYAEPLITGPEYATPIIMDSGHPTTSVGQPS 120
QY 577 STFKATGNOPPLVGYNTLLSRTDSCSSAQAOYDTPKAGKPLPAPDELVYQVPOSTQE 636
DB 121 STFKATGNOPPLVGYNTLLSRTDSCSSAQAOYDTPKAGKPLPAPDELVYQVPOSTQE 180
QY 637 VSGAGRGDCDVFKEIL 653
DB 181 VSGAGRGDCDVFKEIL 197

RESULT 4
ID AAE22721 standard; Protein; 365 AA.
XX
AC AAE22721;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy2 mature protein.
XX
KW Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neurotropic; neuroprotective; vulnerability; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive.
XX
OS Homo sapiens.
XX
PN WO200222815-A1.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28488.
XX
PR 11-SEP-2000; 2000US-0659671.
PR 06-SEP-2001; 2001US-317902P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT;
XX
DR WPI; 2002-393966/42.
XX
PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects -
XX

PS Disclosure; Page 131-132; 152pp; English.

XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.
CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat
CC platelet disorders e.g. thrombocytopenia, plastic anaemia and paroxysmal
CC nocturnal haemoglobinuria and is used in nerve tissue growth or
CC regeneration, in wound healing, tissue repair and replacement and in
CC healing of bones, incisions and ulcers. Compositions comprising the
CC sequences of the invention are useful for treating diseases of peripheral
CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
CC stroke, ulcers, immune deficiencies and immune disorders, infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy2
CC mature protein.

XX Sequence 365 AA;

QY Query Match 24.0%; Score 157; DB 23; Length 365;
Db Best Local Similarity 100.0%; Pred. No. 6.3e-145;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GKCYGLQMHNSIESKNEITLLFMSCGHVSGRGLASVVDKQDITCLDTASNFLE 82
Db 79 GKCYGLQMHNSIESKNEITLLFMSCGHVSGRGLASVVDKQDITCLDTASNFLE 138

QY 83 PEFKYPAGCLLPFAEISGTHPHGRDSSPLCMAGVHAGVYVNTLGGQISVWISKIPY 142
Db 139 PEFKYPAGCLLPFAEISGTHPHGRDSSPLCMAGVHAGVYVNTLGGQISVWISKIPY 198

QY 143 YESSLANNVTSVGHLSLFTFTKSGCYGLMGESG 179
Db 199 YESSLANNVTSVGHLSLFTFTKSGCYGLMGESG 235

RESULT 5
AAE22716
XX ID AAE22716 standard; Protein; 385 AA.
XX AC AAE22716;
XX DT 09-AUG-2002 (first entry)
XX DE Human neuropilin-Hy2 protein.

XX Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW nontropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive; chromosome 6q21.

XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH 1..20
FT Peptide /label= Signal_peptide
FT Protein 61..385
FT /note= "Human mature neuropilin-Hy2 protein".
XX WO200222815-A1.
PN
XX 21-MAR-2002.
PD
XX 12-SEP-2001; 2001WO-US28488.
PF
XX 11-SEP-2000; 2000US-0659671.
PR
XX 06-SEP-2001; 2001US-317902P.
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT;
PI WPI; 2002-393966/42.
DR N-PSDB; AAD35994.
XX
XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects
XX
XX Claim 3; Page 128-130; 152pp; English.

XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases.
CC like Alzheimer's disease and for treating learning and memory disorders.
CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat
CC platelet disorders e.g. thrombocytopenia, plastic anaemia and paroxysmal
CC nocturnal haemoglobinuria and is used in nerve tissue growth or
CC regeneration, in wound healing, tissue repair and replacement and in
CC healing of bones, incisions and ulcers. Compositions comprising the
CC sequences of the invention are useful for treating diseases of peripheral
CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
CC stroke, ulcers, immune deficiencies and immune disorders, infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy2
CC protein. Neuropilin-Hy2 gene is located on chromosome 6q21.

XX Sequence 385 AA;

QY Query Match 24.0%; Score 157; DB 23; Length 385;
Db Best Local Similarity 100.0%; Pred. No. 6.6e-145;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GKCYGLQMHNSIESKNEITLLFMSCGHVSGRGLASVVDKQDITCLDTASNFLE 82
Db 99 GKCYGLQMHNSIESKNEITLLFMSCGHVSGRGLASVVDKQDITCLDTASNFLE 158

QY 83 PEFKYPAGCLLPFAEISGTHPHGRDSSPLCMAGVHAGVYVNTLGGQISVWISKIPY 142
Db 159 PEFKYPAGCLLPFAEISGTHPHGRDSSPLCMAGVHAGVYVNTLGGQISVWISKIPY 218

QY 143 YESSLANNTSVVGHLSLSTFTFKSGCYGTLMESG 179
DB 219 YESSLANNTSVVGHLSLSTFTFKSGCYGTLMESG 255

RESULT 6
AAU79460
ID AAU79460 standard; Protein; 385 AA.
AC AAU79460;
DT 15-JUL-2002 (first entry)
DE Human Neuropilin-Hy2.
XX
XX Human; neuropilin-Hy2; chromosome 6q21; neuronal growth;
KW nerve regeneration; neurodegenerative disease; learning disorder;
KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
KW organ growth; nervous system lesion; cancer; cell proliferation;
KW cell differentiation; stem cell growth factor activity;
KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
KW platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
KW reperfusion; food supplement; DNA microarray.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 1..20
FT /label= Signal_peptide
FT 21..385
FT /label= Mature_neuropilin_Hy2
XX
XX WO200222780-A2.
XX
XX 21-MAR-2002.
XX
XX 11-SEP-2001; 2001WO-US28590.
XX
XX 11-SEP-2000; 2000US-0659671.
PR 06-SEP-2001; 2001US-0659671.
XX
XX (TANG/) TANG T Y.
XX
XX Tang TY;
XX
XX WPI: 2002-351861/38.
DR N-PSDB; ABK49567.
XX
XX New neuropilin-like polypeptides for diagnosing, preventing and
PT treating neurological conditions and disorders, cancers, and for
PT inducing angiogenesis and neovascularisation.
XX
XX Claim 3; Page 123-125; 144pp; English.

The invention relates to an isolated neuropilin-like polypeptide
CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
CC encoding the proteins and the coding regions of the cDNAs. Also included
CC is a nucleic acid array comprising the cDNAs attached to a surface used
CC for detecting full-matches or mismatches to the cDNAs. The genes
CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
CC The nucleic acid array is useful for detecting full-matches or mismatches
CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
CC in modulating neuronal growth, regenerative capacity, treating
CC neurodegenerative diseases, learning and memory disorders, diagnosing and
CC mapping genetic neuronal defects and degenerative diseases like
CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
CC and organ growth and development (e.g. the heart). The nervous system
CC disorders include lesions of central or peripheral nervous systems,
CC including traumatic lesions, ischaemic lesions, infectious lesions,
CC degenerative lesions, lesions associated with nutritional diseases or
CC disorders, neurological lesions, and lesions caused by toxic substances.
CC The neuropilin-like proteins and cDNAs are also useful as markers for

CC cancers. The neuropilin-like proteins are useful for regulating cell
CC proliferation, cell differentiation, stem cell growth factor activity,
CC for inducing proliferation of neural cells, regeneration of nerve and
CC brain tissue, for treatment of central and peripheral nervous system
CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopaenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues and as a food supplement or molecular
CC weight marker. The cDNAs are useful in gene identification, genome
CC mapping, transgenics, as hybridisation probes, for primer design, for
CC gene chips and as a DNA antigen. The present sequence represents
CC neuropilin-Hy2.
XX
SQ Sequence 385 AA;
Query Match 24.0%; Score 157; DB 23; Length 385;
Best Local Similarity 100.0%; Pred. No. 6.6e-145; Indels 0; Gaps 0;
Matches 157; Conservative 0; Mismatches 0;
QY 23 GKCYGLQLQNMHSIESKGNITLLFMSGIRHVSGRGLASVYIDKDLITCLDTASNFLE 82
DB 99 GKCYGLQLQNMHSIESKGNITLLFMSGIRHVSGRGLASVYIDKDLITCLDTASNFLE 158
QY 83 PEFKYCPAGCLLPFAEISGTIPHGYSRDSPLCMAGVHAGVWSNTLGGQISVVISKIPY 142
DB 159 PEFKYCPAGCLLPFAEISGTIPHGYSRDSPLCMAGVHAGVWSNTLGGQISVVISKIPY 218
QY 143 YESSLANNTSVVGHLSLSTFTFKSGCYGTLMESG 179
DB 219 YESSLANNTSVVGHLSLSTFTFKSGCYGTLMESG 255

RESULT 7
ABB97386
ID ABB97386 standard; Protein; 385 AA.
XX
XX ABB97386;
DT 27-JUN-2002 (first entry)
XX
XX Novel human protein SEQ ID NO: 654.
XX
XX Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; ES;
KW expressed sequence tag.
XX
XX Homo sapiens.
XX
XX WO200222660-A2.
XX
XX 21-MAR-2002.
XX
XX 10-SEP-2001; 2001WO-US26015.
XX
XX 11-SEP-2000; 2000US-0659671.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
XX WPI: 2002-292408/33.
DR N-PSDB; ABN32572.
XX
XX An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
XX Claim 20; SEQ ID NO 654; 509pp; English.
XX
XX The present invention provides the protein and coding sequences of 444

CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a protein of the invention.
 XX
 XX Sequence 385 AA;
 SQ
 Query Match 24.0%; Score 157; DB 23; Length 385;
 Best Local Similarity 100.0%; Pred. No. 6.6e-145;
 Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 23 GKYGGLGNHSTESKNEITLFGSIHVSGRFLASYSVIDKQDLITCLDTASNFLE 82
 DB 99 GKYGGLGNHSTESKNEITLFGSIHVSGRFLASYSVIDKQDLITCLDTASNFLE 158
 QY 83 PEFKYCPAGCLLPFAEISGTPHGYRDSPLCHAGVHAGVVSNTLGGQISVWISKGPY 142
 DB 159 PEFKYCPAGCLLPFAEISGTPHGYRDSPLCHAGVHAGVVSNTLGGQISVWISKGPY 218
 QY 143 YESSLANNVTSVGHLSLSTFTKTSCTGYTLGMSG 179
 DB 219 YESSLANNVTSVGHLSLSTFTKTSCTGYTLGMSG 255
 RESULT 8
 AAEE22720
 ID AAE22720 standard; peptide; 20 AA.
 AC AAE22720;
 XX
 XX 09-AUG-2002 (first entry)
 DT
 XX Human neuropilin-Hy2 signal peptide.
 DE
 XX Human: neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopoiesis; memory; platelet; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW neutropenic; neuroprotective; vulvovaginitis; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquiliser; virucide; antibacterial; cytostatic;
 KW immunosuppressive.
 XX
 XX Homo sapiens.
 OS
 XX WO200222815-A1.
 PN
 XX 21-MAR-2002.
 PD
 XX 12-SEP-2001; 2001WO-US28488.
 PF
 XX 11-SEP-2000; 2000US-0659671.
 PR
 XX 06-SEP-2001; 2001US-317902P.
 PD
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT;
 PI
 XX WPI; 2002-393966/42.
 DR
 XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's

PT disease, and for diagnosing and mapping genetic neuronal defects
 XX
 XX Disclosure: Page 130-131; 152pp; English.
 XX
 CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy2
 CC signal peptide.
 XX
 XX Sequence 20 AA;
 SQ
 Query Match 3.1%; Score 20; DB 23; Length 20;
 Best Local Similarity 100.0%; Pred. No. 7.6e-12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MPLFLLLLLLLLLLEDAGA 20
 DB 1 MPLFLLLLLLLLLLEDAGA 20
 RESULT 9
 AAU74424
 ID AAU74424 standard; Protein; 629 AA.
 AC AAU74424;
 XX
 XX 23-APR-2002 (first entry)
 DT
 XX Mouse protein sequence #3, related to isolation of genes within SLE-1B.
 DE
 XX Mouse; systemic lupus erythematosus 1B; SLE-1B; dermatological;
 KW antiinflammatory; immunosuppressive; systemic autoimmune disorder;
 KW signalling lymphocyte activation molecule; SLAM; lymphocyte antigen 9;
 KW Ly-9; 2B4; natural killer cell receptor; CD48; CD84; LY108; CS1; DEND;
 KW NIT1; upstream transcription factor 1; USF 1; GOLGA4; immune tolerance.
 XX
 XX Mus musculus.
 OS
 XX WO2001188200-A2.
 PN
 XX 22-NOV-2001.
 PD
 XX 17-MAY-2001; 2001WO-US16051.
 PF
 XX 17-MAY-2000; 2000US-204963P.
 PR
 XX 21-SEP-2000; 2000US-234457P.
 PD
 XX (TEXA) UNIV TEXAS SYSTEM.
 PA
 XX Wakeland EK, Wandstrat A, Morel L;
 PI

XX WPI; 2002-066695/09.
 DR N-PSDB; AAS20419.
 XX
 PT Screening for susceptibility to systemic autoimmune disorder by
 PT screening for a mutation within the systemic lupus erythematosus-1B
 PT loci
 XX
 PS Disclosure: Page 115-117; 128pp; English.
 XX
 CC The present invention relates to a new method for screening for
 CC susceptibility to a systemic autoimmune disorder. The method comprises
 CC screening for at least one mutation within the systemic lupus
 CC erythematosus (SLE)-1B loci. Screening for susceptibility to autoimmune
 CC disorders such as systemic lupus erythematosus involves screening for at
 CC least one mutation in a gene or genes with the SLE-1B loci such as a gene
 CC encoding signalling lymphocyte activation molecule (SLAM), lymphocyte
 CC antigen (Ly)-9, 2B4 (a natural killer cell receptor), CD48, CD84, LY108,
 CC CS1, DEDD, NTR1, upstream transcription factor (USF)1, GOLGA4. The method
 CC of the invention is useful for treating SLE and involves administering
 CC a construct comprising a wild-type sequence encoding any one of the
 CC above mentioned genes. Gene therapy also involves the use of antisense
 CC constructs or ribozymes directed against the above mentioned genes for
 CC treating SLE.
 CC Note: The present protein sequence represents a mouse protein of the
 CC invention that relates to the isolation of genes within SLE-1B that
 CC mediate a break in immune tolerance. This sequence is shown in the
 CC sequence listing but is not further defined in the specification.
 XX
 SQ Sequence 629 AA;
 Query Match 1.5%; Score 10; DB 23; Length 629;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 LLLLLLLLLL 14
 Db 435 LLLLLLLLLL 444
 RESULT 10
 AAW70900
 ID AAW70900 standard; Protein; 677 AA.
 XX
 AC AAW70900;
 XX
 DT 14-OCT-1998 (first entry)
 XX
 DE Human beta galactosidase protein.
 XX
 KW Retroviral vector; gene delivery vehicle; expression;
 KW non-immunogenic selectable marker; gene therapy; activation;
 KW human; beta galactosidase.
 XX
 OS Homo sapiens.
 XX
 PN WO9830709-A2.
 XX
 PD 16-JUL-1998.
 XX
 PF 14-JAN-1998; 98WO-US00715.
 XX
 PR 13-JAN-1998; 98US-0038339.
 PR 14-JAN-1997; 97US-0035473.
 PR 27-FEB-1997; 97US-0038339.
 XX
 PA (CHAD/) CHADA S.
 PA (JOLLY/) JOLLY D J.
 PA (MOOR/) MOORE M D.
 XX
 PI Chada S, Jolly DJ, Moore MD;
 XX
 XX WPI; 1998-399153/34.
 DR

DR N-PSDB; AAV42728.
 XX
 PT Non-immunogenic pro:drug activating enzyme(s) and selectable
 PT marker(s) - are used in gene therapy for the treatment of a wide
 PT variety of disorder(s)
 XX
 PS Example 2; Fig 5A-B; 121pp; English.
 XX
 CC The present sequence represents human beta galactosidase and is used as
 CC a marker in the retroviral vector of the invention. The specification
 CC describes a gene delivery vehicle which directs expression of a
 CC non-immunogenic selectable marker or molecule which is capable of
 CC activating a previously inactive compound. Vectors expressing the
 CC markers can be used to deliver a molecule into a target area where it
 CC may cause the activation of a previously inactive substance.
 XX
 SQ Sequence 677 AA;
 Query Match 1.5%; Score 10; DB 19; Length 677;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 LLLLLLLLLL 14
 Db 9 LLLLLLLLLL 18
 RESULT 11
 AAB56675
 ID AAB56675 standard; Protein; 696 AA.
 XX
 AC AAB56675;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Human prostate cancer antigen protein sequence SEQ ID NO:1253.
 XX
 KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
 KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
 KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200055174-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05988.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-587513/35.
 DR N-PSDB; AAF15878.
 XX
 PT Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -
 XX
 PS Claim 11; Page 1672-1675; 2338pp; English.
 XX
 CC AAF15566 to AAF16505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytostatic,
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,

CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
 CC (e.g. diabetes), haematopoietic disorders, neural disorders
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
 CC (e.g. osteoporosis, arthritis), ASG63326-ABG65518 represent albumin
 CC fusion proteins of the invention.

XX Sequence 89 AA;
 SQ Query Match 1.4%; Score 9; DB 23; Length 89;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLL 13
 Db 18 LLLLLVLL 26

RESULT 14
 AAEE21629
 ID AAEE21629 standard; Protein; 89 AA.
 AC AAEE21629;
 XX
 XX 16-JUL-2002 (first entry)
 DT
 DE Human gene 1 encoded secreted protein HCEFB80, SEQ ID NO:40.
 KW Human; secreted protein; immunodeficiency; inflammation; septic shock;
 KW reperfusion injury; inflammatory bowel disease; autoimmune disorder;
 KW rheumatoid arthritis; gastrointestinal disorder; respiratory disorder;
 KW asthma; allergy; cancer; multiple sclerosis; cardiovascular disorder;
 KW neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
 KW atherosclerosis; acquired immune deficiency syndrome; prion disease;
 KW AIDS-related dementia; hepatitis; blood-related disorder; thrombosis;
 KW hyperproliferative disorder; renal disorder; acute glomerulonephritis;
 KW endocrine disorder; Addison's disease; reproductive system disorder;
 KW endometriosis; infectious disease; pancreatic disorder; gene therapy;
 KW vaccine.

XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..35
 FT /label= Signal_peptide
 FT 36..89
 FT /label= Mature_secretured_protein
 FT
 XX WO200216576-A1.
 PN
 XX
 PD 28-FEB-2002.
 XX
 PF 17-JAN-2001; 2001WO-US01396.
 XX
 PR 18-AUG-2000; 2000US-226280P.
 PR 21-DEC-2000; 2000US-256968P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
 PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
 PI NI J;
 XX WPI: 2002-315464/35.
 DR N-PSDB; AAD34068.
 DR
 XX Human secreted polypeptides and polynucleotides for diagnosing,
 PT prognosing, preventing and treating immune, hyperproliferative, liver,
 PT kidney, and reproductive disorders and for identifying their modulators
 PT for therapeutic use
 XX
 PS Claim 11; Page 425; 462pp; English.
 XX

CC AAD34068-AAD34096 represent cDNAs corresponding to 11 human secreted
 CC protein genes, and AAEE21629-AAEE21657 represent the proteins they encode.
 CC AAEE21658-AAEE21663 represent human secreted protein fragments. The genes
 CC and their corresponding secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 11 genes, based on the tissues in which they are most highly expressed.
 CC Sequences of the invention are useful in treating, preventing, diagnosing
 CC and/or prognosing immunodeficiencies, inflammatory conditions (e.g.,
 CC septic shock, reperfusion injury, inflammatory bowel disease), autoimmune
 CC disorders (e.g., rheumatoid arthritis), gastrointestinal disorders,
 CC respiratory disorders (e.g., asthma and allergy), cancers (e.g. liver
 CC and breast), central nervous system (CNS) disorders (e.g., multiple
 CC sclerosis), cardiovascular disorders (e.g., atherosclerosis), neuro-
 CC degenerative disorders (e.g., Parkinson's disease, Alzheimer's disease,
 CC AIDS-related dementia, prion disease) and many additional diseases,
 CC conditions and disorders that are characterised by inflammation (e.g.,
 CC hepatitis). They are also useful for treating blood-related disorder
 CC (e.g., thrombosis), hyperproliferative disorders, renal disorders (e.g.,
 CC acute glomerulonephritis), endocrine disorders (e.g., Addison's disease),
 CC reproductive system disorders (e.g., endometriosis), infectious diseases
 CC and pancreatic disorders. They are also useful as vaccine adjuvants that
 CC enhance immune responsiveness to an antigen and as adjuvants to enhance
 CC tumour-specific immune responses, anti-viral, anti-bacterial, anti-fungal
 CC and anti-parasitic immune responses. Secreted proteins of the invention
 CC are used for enhancing or inhibiting complement mediated cell lysis, for
 CC stimulating wound and tissue repair, angiogenesis and the repair of
 CC vascular or lymphatic diseases or disorders. They stimulate neuronal
 CC growth and are used to treat, prevent and/or diagnose neuronal damage,
 CC to prevent hair loss, to stimulate keratinocyte growth, to modulate
 CC mammalian characteristics such as body height, weight, hair colour and
 CC to increase or decrease storage capabilities, fat content, lipid,
 CC protein, vitamins, carbohydrate, minerals, cofactors or other nutritional
 CC components. Sequences of the invention are also useful in gene therapy.
 CC The present sequence represents a human secreted protein of the
 CC invention.

XX
 SQ Sequence 89 AA;
 Query Match 1.4%; Score 9; DB 23; Length 89;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLL 13
 Db 18 LLLLLVLL 26

RESULT 15
 AAU83189
 ID AAU83189 standard; Protein; 150 AA.
 XX
 AC AAU83189;
 XX
 XX 08-MAY-2002 (first entry)
 DT
 XX
 DE Novel secreted protein Z887280G5P.
 XX
 KW Protein secretion; mammalian secreted polypeptide; MSP.
 KW
 XX Homo sapiens.
 OS
 XX WO200202621-A2.
 PN
 XX 10-JAN-2002.
 PD
 XX 28-JUN-2001; 2001WO-US20638.
 PF
 XX 30-JUN-2000; 2000US-215446P.
 PR
 XX (ZYMO) ZYMOGENETICS INC.
 PA

```

XX Sheppard PO, Presnell SR;
PI WPI; 2002-147999/19.
XX N-PSDB; ABK33104.
DR
XX Novel isolated mammalian secreted polypeptide useful in therapeutic and
PT diagnostic methods, to direct secretion of other proteins of interest
PT from host cell, as educational tools, and as laboratory practicum kits
PT
XX
XX Claim 12: Page 294; 397pp; English.
XX The invention describes an isolated mammalian secreted polypeptide (MSP)
CC (1). (1) is useful to direct the secretion of other proteins of interest
CC from a host cell, to monitor secretion of proteins, to degenerate
CC sequences comprising all nucleotide sequences encoding a particular
CC polypeptide, to screen for cell metabolism effecting receptors, for
CC identifying new target receptors and drug design, for identifying, for
CC protein purification, for determining the weight of expressed MSP,
CC polypeptides as a ratio to total protein expressed, for identifying
CC peptide cleavage sites, for coupling amino and carboxy terminal tags, for
CC amino acid sequence analysis, for monitoring biological activities of the
CC protein in vitro and in vivo, and to teach analytical skills and as
CC reagents for the study of cells, receptors, and other binding molecules.
CC The polynucleotide is useful for radiation hybrid mapping, and somatic
CC cell genetic technique developed for constructing high-resolution,
CC contiguous maps of mammalian chromosomes. Reagents disclosed in the
CC invention may be used to detect metabolic abnormalities characterised by
CC over or under production of the protein. This is the amino acid sequence
CC of a mammalian secreted polypeptide, described in the method of the
CC invention.
XX
XX SQ Sequence 150 AA;
Query Match 1.4%; Score 9; DB 23; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 LLLVLLLLL 15
DB 6 LLLVLLLLL 14
Search completed: January 21, 2003, 09:57:15
Job time : 26.5152 secs

```


Alignment Scores:

Pred. No.: 2,72e-26 Length: 432
 Score: 412.00 Matches: 145
 Percent Similarity: 33.08% Conservative: 71
 Best Local Similarity: 22.21% Mismatches: 141
 Query Match: 9.94% Indels: 296
 DB: 11 Gaps: 19

US-10-060-830-1 (1-2280) x Q8R327 (1-432)

```

QY 364 GAATAGCAATATCTGCTGGGTGCAATGAACCAATCAATGAATCAAAAGGC 423
Db 27 GluLeuGlyProTyrCysGly---SerTrpAlaValProLysGluLeuArgLeuAsnSer 45
QY 424 AATGAATACATCTGCTGTCATGAGTGGATCCATCTTCTGGACGCGGATTTTGGCC 483
Db 46 AsnGluValThrValLeuPheLysSerGlySerHisLeuSerGlyArgGlyPheLeuLeu 65
QY 484 TCATCTCTGTATAGATAACAAGATCAATTAATCTGTTTGGACACTGCATCCAAATTT 543
Db 66 ThrTyrAlaSerSerAspHisProAspLeuIleThrCysLeuGluArgGlySerHisTyr 85
QY 544 TTGGAACCTGAGTTCAGTAAGTACTGCCAGCTGTTGCTGCTCTCTTTGCTGAGATA 603
Db 86 PheGluGluLysTyrSerLysPheCysProAlaGlyCysArgAspIleAlaGlyAspIle 105
QY 604 TCTGGAACAATTCCTCATGATATAGAGATTCCTCGCCATTTGTCATGGCTGGTGCAT 663
Db 106 SerGlyAsnThrLysAspGlyTyrArgAspThrSerLeuLeuCysLysAlaAlaIleHis 125
QY 664 CGAGGAGTAGTGTCAACACGCTGGGGCCCAATCAGTGTGTGAATAGTAAGGTATT 723
Db 126 AlaGlyIleIleThrAspGluLeuGlyHisIleAsnLeuLeuGlnSerLysGlyIle 145
QY 724 CCTATTATGAAGTCTTTGGCTAAACAACGTCACATCTGTGGTGGGACACTTATCACA 783
Db 146 SerHisTyrGluGlyLeuLeuAlaAsnGlyValLeuSerArgHisGlySerLeuSerGlu 165
QY 784 AGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGGTGTG 843
Db 166 LysArgPheLeuPheThr--- 172
QY 844 ATCGGGATCTCAATACAGCATCATCTGTGCTGGAGTGGACTGACCACACAGGCCAA 903
Db 172 --- 172
QY 904 GAGACAGTTGGAACCCCAAAAGCCAGCGTGAATAAACCTGGACGCCCTTGGCGTCT 963
Db 172 --- 172
QY 964 TTGGCCACTGATGAATACCAGTGGTTACAAATAGATTTGAATAAGGAAAGAAATAACA 1023
Db 172 --- 172
QY 1024 GGCATTATACCACTGGATCCACCATGTTGGAGCACAAATTACTATGCTCTGCTGCACAGA 1083
Db 172 --- 172
QY 1084 ATCCTGTACAGTGATGGGAGAAATGAGCTGTGTACAGAGACGCTGGTGTGGAGCAA 1143
Db 173 --- 175
QY 1144 GATAAGATATTTCAAGGAACAAGATTTATCACCAGGATGTGCGTAATAAATTTTGGCCA 1203
Db 175 --- 175
QY 1204 CCAATTATTGCAGTTTTATTAGTGAATCCTACCAATGCCACAGAGAAATTCGCCATG 1263
Db 175 --- 175
QY 1264 AAAATGGAGCTGCTGGATGTCAGTTTATTCCTAAAGGTGCTCTCCAAAACCTTACTCAA 1323
Db 175 --- 175

```

```

QY 1324 CTCCACCTCTCGGACACGAATGACCTCAAAAACACTACAGCCCTCCAAAANTAGCC 1383
Db 175 --- 175
QY 1384 AAAGGTGTCGCCCAAAATTTACGCCACCACTACACCTCGCAGTAGCAATGAATTTCT 1443
Db 175 --- 175
QY 1444 GCACAGACAGAACAAACAACTGCCAGTCTCTGATATACAGAAATACCGTAACCTCAAAT 1503
Db 176 --- 176
QY 1504 GTACCAAAAGATAGCGCTGGTGGCGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1563
Db 177 IleThr---ThrValAlaIleProSerValIle-----PheIleAlaLeuLeuLeuThr 193
QY 1564 ACTCTCATCTCATATTAGTGTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1623
Db 194 GlyMetGlyIlePheAlaIleCys-----ArgLysArgLysLysLys 207
QY 1624 GGCACCTATGACTTACCTTAC-----TGGACCGGGCAGGTGTGTGGAAGGA 1671
Db 208 GlyAsn-----ProTyrValSerAlaAspAlaGlnLysThrGlyCysTrpLysGln 224
QY 1672 ATGAGCAGTTCCTCTGCAAAAGCAGTGACCATGAGGAACCCCA---GTTGCTCT 1728
Db 225 IleLysTyr-----ProPheAlaArgHis 232
QY 1729 AGCAGCAGCGAA-----GTTAATCACCTGAGTCCAAGAGAAGTCAC---ACA 1773
Db 233 GlnSerThrGluPheThrIleSerTyrAspAsnGluLysGluMetThrGlnLysLeuAsp 252
QY 1774 GTGCTGAGGCTGACTCTGACAGATATGCTCAGCCACTGTGAGGAGGAATGTGTGTACA 1833
Db 253 LeuIleThrSerAspMetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrVal 272
QY 1834 CTTCATCAAGATCTACCTTTAAACCA-----GAGAA----- 1866
Db 273 AlaArgLysGlySerThrPheArgProMetAspThrAspThrGluGluValArgValAsn 292
QY 1867 GGAAGAAGCAGCGCTATGCAGACCTAGACTCTTACCACTCACCAGCGCAGCAAGTTTAT 1926
Db 293 ThrGluAlaSerGlyHis-----TyrAspCysProHisArgProGlyArg 307
QY 1927 CATCCCTATGCTGAACCACTCCCAATACGGGGCTGATGTGCAACCCCAATCATC--- 1983
Db 308 HisGluTyrAlaLeuProLeuThrHisSerGluProGluTyrAlaThrProIleValGlu 327
QY 1984 --- 2016
Db 328 ArgHisLeuLeuArgAlaHisThrPheSerThrGlnSerGlyTyrArgValProGly--- 346
QY 2017 GGTCAAGCTCCACATCCACTTTCAGGCTACGGGGAACCAACCTCCGCCACTAGTGGGA 2076
Db 347 ProArgProThrHisLysHisSerHisSerGlyGlyPheProProAlaThrGlyAla 366
QY 2077 ACTTACATACACTTCTCTCCAGGACTGACAGCTCTCTCAGCCAGCCAGCCAGTATGAT 2136
Db 367 ThrGlnIleGluSerTyrGlnArgProAlaSerProLysProValGlyGlyTyrAsp 386
QY 2137 ACCCGAAGCTGGGAAGCCAGGTCTACTGCCCCAGCAGCAATTTGGTGACCAGTGCCA 2196
Db 387 LysPro---AlaAlaSerSerPheLeuAspSerArgAsp-----Pro 399
QY 2197 CAGACCACACAAGAGTATCAGGACGAGGAGGATGGG 2235
Db 400 AlaSerGlnSerGlnMetThrSerGlyGlyAspGly 412

```

RESULT 13

Q96190

ID Q96190

AC Q96190;

PRELIMINARY;

PRT: 609 AA.

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	417.5	10.1	925	4	US-09-116-473-2	Sequence 2, Appl
2	414.5	10.0	901	3	US-08-936-135-22	Sequence 22, Appl
3	414.5	10.0	906	3	US-08-936-135-24	Sequence 24, Appl
4	414.5	10.0	909	3	US-08-936-135-8	Sequence 8, Appl
5	412.5	10.0	909	3	US-08-936-135-10	Sequence 10, Appl
6	412.5	10.0	914	3	US-08-936-135-12	Sequence 12, Appl
7	412.5	10.0	926	3	US-08-936-135-14	Sequence 14, Appl
8	412.5	10.0	931	3	US-08-936-135-16	Sequence 16, Appl
9	398.5	9.6	909	3	US-08-936-135-18	Sequence 18, Appl
10	398.5	9.6	926	3	US-08-936-135-20	Sequence 20, Appl
11	392.5	9.5	923	3	US-08-936-135-6	Sequence 6, Appl
12	388.5	9.4	922	4	US-09-116-473-4	Sequence 4, Appl

; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 925 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-116-473-2

Alignment Scores:

Pred. No.: 4,53e-30 Length: 925
 Score: 417.50 Matches: 200
 Percent Similarity: 36.30% Conservative: 110
 Best Local Similarity: 23.42% Mismatches: 348
 Query Match: 10.08% Indels: 196
 DB: 4 Gaps: 34

US-10-060-830-1 (1-2280) x US-09-116-473-2 (1-925)

QY 76 ATGCCTCTGTCCTCTCTTACTTGTCTGCTCTGCTGCTC-----GAGGAGCCT 129
 Db 1 MetAspMetPheProLeuThrTrpPheLeuAlaLeuTyrPheSerGlyHisLysVal 20
 QY 130 GGACCCAGCAGGTGATGGATGGACACACTTACTAGGCCCTGAGAGTGGACCCCT 189
 Db 21 ArgSerGlnGlnAspProProCysGlyGlyArgLeuAsnSerLysAspAlaGlyTyrIle 40
 QY 190 ACATCCATAACTACCCACAGACCTATCCCAACAGCAGCTGTTGTGAATGGAGATCCGT 249
 Db 41 ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluTrpValTyr 60
 QY 250 GTA---AAGATGGAGAGAGTTCGCATCAAAATTTGGT---GACTTTGACATTGAAGAT 303
 Db 61 AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGluLys 80
 QY 304 TCTGATCTCTCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 363
 Db 81 HisAsp---CysLysTyrAspPheIleGluIleArgAspGlyAspSerGluSerAlaAsp 99
 QY 364 GAAATAGCAAACTACTGTGGTGGGTGGGTGCAATGAAC-----CATTCAAATGAA 414
 Db 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProThrIleIle 114
 QY 415 TCAAAGCAATGAATACATNTGCTGTTTCATGAGTGGAAATCCATGTTTTCGACGGGA 474
 Db 115 SerSerGlySerValLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly 134
 QY 475 TTTTGGCCCTCATCTCTGTATAGATAAACAAGATCAATTAATTAATTAATTAATTAAT 534
 Db 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
 QY 535 TCCAAATTTTGGAACTGAGTTCAGTAAGTACGTCGCGAGTGTTGCTGCTTCCTTTT 594
 Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
 QY 595 GCTGAGATATCTGGAACAATTCCTCATGGATATAGAGATCTCTGCCATTGTCATGGCT 654
 Db 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
 QY 655 GGTGTGATGCAGAGTAGTCTCAACACAGCTGGTGGCGGCCCAATCACT----- 702
 Db 182 LysProArgMetGluIleIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu 201
 QY 703 -----GTTGTAATTAAGTAAGTATTCCG--- 726
 Db 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis 221
 QY 727 -----TATTATGAAGTCTTTTGGCTAACACACGACATCTGTG 765
 Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSerSer 241
 QY 766 GTGGACACATTTATCTACAGTCTTTTACATTTAAGACAAGT----- 807
 Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258

QY 808 -----GGATGTTAT 816
 Db 259 GlyPheSerAlaArgTyrTyrLeuValHisGlnGluProProGluAsnPheGlnCysAsn 278
 QY 817 GGAACACTGGGATGGAGTCTGTTGTCGATCGCGATCTCTCAATAACAGCATCATCTGTG 876
 Db 279 AlaProLeuGlyMetGluSerGlyArgIleValAsnGluGlnIleSerAlaSerSer--- 297
 QY 877 CTGGAGTGGACTGACACACAGAGGCGAAGAACAGTGTGAAACCCCAAAAAGCCAGGCTG 936
 Db 298 -----ThrPheSerAspGlyArgTyrThrProGlnGlnSerArgLeu 311
 QY 937 AAAAAACCTGGACCGCTTGGCTGCTTTGCCACTGATGAATACCACTGTTTACAATA 996
 Db 312 HisGlyAspAspAsnGlyTrpThrProAsnValAspSerAsnLysGluTyrLeuGlnVal 331
 QY 997 GATTTGNAATAGGNAAGAAATAACAGGCATTATATACCACTGGA-----TCCACCATG 1050
 Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
 QY 1051 GTGGAGCACAAATTAATGTCCTGCTACAGAACTCTGACAGTATGATGGCGACAAA 1110
 Db 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyLysAsp 371
 QY 1111 TGGACTGTGTACAGAGAGCCTGCTGTGGACAGATAGATATTTTCAAGGAACAAGAT 1170
 Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysValPheGlnAlaAsnAsp 389
 QY 1171 TATCACAGGATGTGCTAATAACTTTTGGCCACCAATTAATGACGCTTTTATTAGAGTG 1230
 Db 390 AlaThrGluLeuValLeuAsnLysLeuHisThrProLeuLeuThrArgPheIleArgIle 409
 QY 1231 AATCTTACCCCAATGGCAGACAGAAATGCCATGAAATGGAGCTCTCGGATGCAGTTT 1290
 Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
 QY 1291 ATTCTTAAAGTGTCTCTCCAAAATTAATCACTCACTCTCTCGCAACAGCAATGAC 1350
 Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
 QY 1351 CTCAAAAACACTACAGCCCTCCAAAATAGCCAAAGTCTGCTGCCCAAAATTT----- 1404
 Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTyr 460
 QY 1405 -----ACGCAACCACTACACCTCGCAGTAGCAATGAATTTCTCGCACACAGACAA 1455
 Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgAsnProGln 480
 QY 1456 CAACAACACTGCCAGTCTGATATCAGAAATTAATACCGTAATCTCAAAATGTAAACCAAGAT 1515
 Db 481 AlaGlnProGlyGluGluTyrLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
 QY 1516 GTA-----GCGTGGCTGCAGTCTTGTCTCCCTGCTGCTG 1548
 Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaMetGluAlaArgAlaPhe 520
 QY 1549 GTCATGTCCTCACTCTCTCATTTCTATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1599
 Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
 QY 1600 -----AGAAACAGAAAGAAAACCTGAAGGCACC-----TATGACTTACCTTACTG 1647
 Db 541 ProArgThrGlnGlnProLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
 QY 1648 GACCGGCGAGTTGGTGGAAAGGAATGAACAGTCTTCTCTGCAAAAGCAGTGGACCAT 1707
 Db 559 -----AspIleArgPhe----- 563
 QY 1708 GAGGAACCCCGAGTCTCCCTATAGCAGCGAGCAAGTAACTACCTGAGTCCCAAGAGATC 1767
 Db 564 GluProValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583

Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSerSer 241
QY 766 GTGGGACACTTATCTACAGTCTTTTACATTTACACAGT----- 807
Db 242 ThrGlyLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 808 -----GGATGTTAT 816
Db 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProGluAsnPheGlnCysAsn 278
QY 817 GGAACCTGGGATGAGTCTGTGTGATCGGGATCCTCAATAACACCATCATCTGTG 876
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer--- 297
QY 877 CTGGAGTGGNACTCACACACAGGCGCAGAGACAGTGTGAAACCCCAAAAGCCAGCGTG 936
Db 298 -----ThrPheSerAspGlyArgTyrPheProGlnGlnSerArgLeu 311
QY 937 AAAAAACCTGGACCGCTTGGCTGCTTTTCCCACTGATGAATACACAGTGTGTACAAATA 996
Db 312 HisGlyAspAsnGlyTyrPheProAsnLeuAspSerAsnLysGluTyrLeuGlnVal 331
QY 997 GATTGTAATAGGAAGAAATAACAGGCAATTATACCACTGGA-----TCCACCATG 1050
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGGACACAATTAATCTGCTGCTACAGAACTCTGTACAGTGTATGGGCAGAAA 1110
Db 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuValSerThrAsnGlyGluAsp 371
QY 1111 TGGACTGTGTACAGAGCCGTGTGTGGACAAGATAAGATATTTCAAGAAACAAAGAT 1170
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsnAsp 389
QY 1171 TATCACAGAGTGTGCTATTAATCTTTTCCACCAATTAATTCACCGTTTTTATAGAGTG 1230
Db 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgile 409
QY 1231 AATCCTCCCAATGAGCAGAGAAATTCGCAATGAGTGTCTGGATGTCAAGTTT 1290
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1291 ATTCCTAAAGTCCTCTCCAAACTTACTCAACCTCCACTCTCTCGGAACAGCAATGAC 1350
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCAAAAACACTACAGCCCTCCAAAAATAGCCAAAGTGTGTCGCCCAAAATTT----- 1404
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTrp 460
QY 1405 -----ACGCAACCACTACAACTCGCAGTAGCAATGAATTTCTGCACAGACAGAA 1455
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgAsnProGln 480
QY 1456 CAACAACCTGCCAGTCTGATATCAGAAATACTACCTGAATCTCAATGTAACCAAGAT 1515
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----CGCTGGCTGCAGTCTTGTCTCCCTGTGTG 1548
Db 501 ValIleIleGlnGlyAlaArgGlyCysSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1549 GTTCATGGTCTCACTACTCTATTCTCATATTAGTGTGTGTGTGGCACTGG----- 1599
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1600 -----AGAACAGAAAGAAACTGAAGCACC-----TATGACTTACTTACTGG 1647
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1648 GACCGGCGAGGTGTGGAAAGGATGAAGCAGTTCTCTCTGCAAAAGCAGTGGACCAT 1707
Db 559 -----AspIleArgPheAspPro----- 565

QY 1708 GAGGAAACCCCACTTCGTCTATAGCAGCAGCAAGTAATACCTGAGTCCACAGAGATC 1767
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1768 ACCACAGTCTCGCAGCTGACTCTGCAGAGTAT-----GCTCAGCCTGTTAGGAGGA 1821
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1822 ATTGTTGTGTACTACTCATCAAGATCTACTTTTAAACCA-----GAAGAAGGA 1869
Db 604 IeuGlyProThrValLysSerGluGluThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAAGAAGCAGGCTAT-----GCAGACCTAGATCTTCACTCA 1908
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerG1 643
QY 1909 CCAGGCGCAGGAAGTTTATCATGCCT-----ATGCTGAACCACTCCCA 1950
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTrpValTyrAspHisAlaL 663
QY 1951 A---TTACGGGCGCTGAGTATGCAACCCCAATCATCATGACATGTCAGGCGCACCCCA-- 2005
Db 663 strPLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 2006 -----CAACTTCAG----- 2014
Db 676 nAspArgThrPheProAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArg1 696
QY 2015 -----TTGTCAGCCTCCACATCCACTTTCAGGCTACGGGCAACCAA 2058
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 2059 CTTCCCTCCACTAGTGGGAATCAATATACACTTCTCTC----- 2096
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2097 -----CAGACTGACAGTGTCT-----CTCAGCCCGCCCGCAG 2130
Db 731 InValValArgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2131 TATGATACCCCGAAAGCTGGGAAGCCAGGCTACTCTGCCACAGACAGAAATGTTGTTACCCAG 2190
Db 751 GlySerGluTrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2191 GTCCACAGACACACAAGTAATCAGAGCAGGAAGGATGGGAATGTGTGTGTT 2247
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGluIleSerIle 786

RESULT 3

US-08-936-135-24
; Sequence 24, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936.135

```

; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-936-135-24

Alignment Scores:
Pred. No.: 8.5e-30 Length: 906
Score: 414.50 Matches: 205
Percent Similarity: 36.59% Conservative: 110
Best Local Similarity: 23.81% Mismatches: 336
Query Match: 10.00% Indels: 210
DB: 3 Gaps: 37

US-10-060-830-1 (1-2280) x US-08-936-135-24 (1-906)
QY 76 ATGCTCTGTCCTCGCTTACTGTCTGCTCTGCTGCTGCTC-----GAGGACGCT 129
DB 1 MetAspMetPheProLeuThrTrpValPheLeuAlaLeuThrPheSerGlyHisGluVal 20
QY 130 GGAGCCAGCAAGGTGATGGATGTGGACACACTGTACTAGCCCTGAGAGTGGACCCCTT 189
DB 21 ArgSerGlnGlnAspProProCysGlyArgProAsnSerLysAspAlaGlyTyrIle 40
QY 190 ACATCCATAAATACCCACAGACTATCCCAACAGCACTGTTTGTGATGGAGATCCGT 249
DB 41 ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluTrpIleValTyr 60
QY 250 GTA---AAGATGGGAGAGAGTTCGCATCAAAATTGGT---GACTTTGACATTGAAGAT 303
DB 61 AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGluLys 80
QY 304 TCTGATTCTTCTCACTTAATTAATCTGAGAAATTAATGGAATTTGGAGTCAGCAGAACT 363
DB 81 HisAsp---CysLysTyrAspPheIleGluLeuArgAspGlyAspSerGluSerAlaAsp 99
QY 364 GAAATAGCAAAATACTGTGCTGGGTGCGAATGAAC-----CATTCATTTGAA 414
DB 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProThrIleIle 114
QY 415 TCAAAAGCAATGAATACATTCCTGCTCATGATGGATCCATGTTCTGGACGCGGA 474
DB 115 SerSerGlySerValLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly 134
QY 475 TTTTGGCCCTCATCTCTGTATAGATAAACAAGATCTAATTAATCTGTTGGACACTGCA 534
DB 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 535 TCCAAATTTTGGACCTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTCTCCCTTTT 594
DB 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
QY 595 GCTGAGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTCGCCATTTGCTGCTGCT 654
DB 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 655 GCTGTGATGTCAGGAGTAGTGTCAACACACAGCTTGGCGGCCCAAAATCAGT----- 702
DB 182 LysProArgMetGluIleIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu 201
QY 703 -----GTTGTAATAGTAAGGTATTCCC----- 726

```

```

DB 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpaspGlyIleProHis 221
QY 727 -----TATTATGAAGATTCTTGGCTTAACACAGTCACATCTGTG 765
DB 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSerSer 241
QY 766 GTGGGACACTTACTACAAAGCTTTTACATTTAAGACAAGT----- 807
DB 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetalAlaValAlaLysAsp 258
QY 808 -----GGATGTTAT 816
DB 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProProGluAsnPheGlnCysAsn 278
QY 817 GGAACACTGGGATGGATGCTGTGTGATCGCGATCCTCAATAACAGCATCATCTGTG 876
DB 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer--- 297
QY 877 CTGGAGTGGACTGACACACAGGCAAGACAGTGGAAACCCAAACCCAGGCTG 936
DB 298 -----ThrPheSerAspGlyArgTrpThrProGlnGlnSerArgLeu 311
QY 937 AAAAAACCTGGACCGCTTGGCTGCTTTGCCACTGATGAATACCACTGGTTACAAATA 996
DB 312 HisGlyAspAspAsnGlyTrpThrProAsnLeuAspSerAsnLysGluTyrLeuGlnVal 331
QY 997 GATTGTAATAAGGAAAAAATAACAGGCATTATAACCACTGGA-----TCCACCATG 1050
DB 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGGAGCACAATTAATGCTGTGCTGCTACAGATCTGTACAGTGTATGGCAGAAA 1110
DB 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1111 TGGACTGTGTACAGAGAGCCCTGGTGGAGCAAGATAAGATATTTCAAGAAACAAAGAT 1170
DB 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsp 389
QY 1171 TATCACCAGGATGTGCGTAATAACTTTTGCACCAATATTGCGACTTTTATTAGAGT 1230
DB 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1231 AATCCTACCCANTGGCAGCAGAAATTTGCCATGAATGGAGCTGCTCGGATGTACGTTT 1290
DB 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1291 ATTCCTAAGGTGCTCTCCAAAACCTTACTCAACCTCCACCTCCTCGGAACAGCAATGAC 1350
DB 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCAAAACACTACAGCCCTCCAAAATAAGCCAAAGTCTGCGCCCAAAATTT----- 1404
DB 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGlyTyrLeuTrp 460
QY 1405 -----ACGCAACACTACACCTCGCAGTAGCAGTAATGATTTCTCGCACACAGAA 1455
DB 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTyrPheProArgAsnProGln 480
QY 1456 CAACAACCTGCCAGTCTGTATATCAGAAATACTACCTTAACCTCAAAATGTAAACCAAGAT 1515
DB 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----CCGCTGGCTGCAGTTCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1548
DB 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1549 GTCATGTCCTCACTACTCTCTCATTTAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1599
DB 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1600 -----AGAAACAGAAAGAAAACCTGAGGCACC-----TATGACTTACCTTACTGCG 1647

```

Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1648 GACCGGGAGGTTGGTGAAGGAATGAAGCAGTTCTTCCTGCAAAAGCAGTGGACCAT 1707
Db 559 -----AspLeuArgGpPheAspPro----- 565
QY 1708 GAGGAACCCAGTTCGCTATAGCAGCAGCAGGAGTAAACACCTGAGTCCAGAGAGATC 1767
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTyrSerProAlaGlyIle 583
QY 1768 ACCACAGTCTGCGAGGCTGACTCTGAGAGATP-----GCTCAGCAGCAGTGTAGGAGGA 1821
Db 584 GlyMetArgLeuValLeuGlyCysAspTyrThrAspSerLysProThrValGluThr 603
QY 1822 ATTGTTGGTACATCTCATCAAGATCTACTTTAAACCA-----GAAGAAGGA 1869
Db 604 LeuGlyProThrValLysSerGluGluThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAGAAGCAGGCTAT-----GCAGACCTAGATCTTACAACTCA 1908
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerGI 643
QY 1909 CCAGGGCAGGAAGTTATCATGCT-----ATGCTGAACCACTCCCA 1950
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTyrValTyrAspHisAlaL 663
QY 1951 A-----TTACGGGGCTGAGTATGCAACCCCAATCATCATGACATGTCAGGGCACCACA-- 2005
Db 663 strLeuArg-----SerThrTrpLysSerSerAlaAsnProAs 676
QY 2006 -----CRACTTCAG----- 2014
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgGI 696
QY 2015 -----TTGGTCAGCCCTCCACATCCACATTTCAAGGCTACGGGGAACCAA 2058
Db 696 uGlyGlnTyrGlyArgLeuLysSerProPro-ValHisLeu-----ProA 711
QY 2059 CTTCCCGCAGTAGGGAACTTACATACACTCTCTC----- 2096
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyHisGlyValAlaLeuG 731
QY 2097 -----CAGGACGACAGCTGCTC-----CTCAGCCCGAGCCGAC 2130
Db 731 lnValValArgGluAlaSerGlnGluSerLysLeuLeuTyrValIle-ArgGluAspGln 750
QY 2131 TATGATACCCGAAAGCTGGAGCCAGGCTACTCTCCCGCAGCAGCAATGGTGTACCAG 2190
Db 751 GlySerGluTyrLysHisGlyArgIleLeuLeuProSerTyrAsp---MetGluTyrGln 769
QY 2191 GTGCCACAGCACACAGAAGATATCAGGACGAGGAGGATGGGGAATGTGATGTT 2247
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGluIleSerIle 786

RESULT 4
US-08-936-135-8
; Sequence 8, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 909 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-8

Alignment Scores:
Pred. No.: 8,51e-30 Length: 909
Score: 414.50 Matches: 205
Percent Similarity: 36.59% Conservative: 110
Best Local Similarity: 23.81% Mismatches: 336
Query Match: 10.00% Indels: 210
DB: 3 Gaps: 37

US-10-060-830-1 (1-2280) x US-08-936-135-8 (1-909)

QY 76 ATGCCCTGTTCTCTGCTCTTACTGTCTCTCTCTCTGCTC-----GAGGACGCT 129
Db 1 MetAspMetPheProLeuThrTrpValPheLeuAlaLeuTyrPheSerGlyHisGluVal 20
QY 130 GGAGCCAGCAGAGGTGATGATGTGGACACACTGTACTAGGCCCTGAGAGTGGACCCCT 189
Db 21 ArgSerGlnGlnAspProProCysGlyGlyArgProAsnSerLysAspAlaGlyTyrIle 40
QY 190 ACATCCATAAATACCACACAGACCTATCCCAACAGCAGCTGTTTGTGATGGAGATCCGT 249
Db 41 ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluTrpIleValTyr 60
QY 250 GTA---AAGATGGGAGAGAGTTCGCATCAAAATTTGCT---GACTTTGACATTGAAGAT 303
Db 61 AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGluLys 80
QY 304 TCTGATCTCTTCTACTTAAATTTACTTGAATTTTATATGGAATTTGGAGTCAGCAGAACT 363
Db 81 HisAsp---CysLysTyrAspPheIleGluLeuArgAspGlyAspSerGluSerAlaAsp 99
QY 364 GAAATAGGCAAAATCTGTGCTGCTGGGTGCAAAATGAAC-----CATTCATTTGAA 414
Db 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProThrIleIle 114
QY 415 TCAAAAGGCAATGAATACACATCTGCTTTCATGATGGAATCCATGTTCTGGAGCGGA 474
Db 115 SerSerGlySerValLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly 134
QY 475 TTTTGGCCCTCATCTCTGTTTATAGATAACAAGATCTAATATTCTGTTTGGACACTGCA 534
Db 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 535 TCCAATTTTGGAAACCTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTGCTTCTCTTTT 594
Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
QY 595 GCTGAGATATCTGGAACAAATTCCTCATCGATATAGAGATTCCTGCCCATTTGTCATGCT 654
Db 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181

QY 655 GGTGTGATGTCAGGAGTGTGTCACACAGCTTGGCGGCCAAATCAGT----- 702
Db 182 LysProArgMetClnuilelleLeuGlnPheLeuThrPheAspLeuHIsAspProLeu 201
QY 703 -----GTTGTAATTAGTAAGAGTATTC----- 726
Db 202 GlnValGlyGluCysPysLysTyrAspTrpLeuAspIleThrPaspGlyIleProHis 221
QY 727 -----TATTATGAAGTTCTTTTACATTAACACGTCATCTGTG 765
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSer 241
QY 766 GTGGACACTTACTACAGTCTTTTACATTAACACGTCATCTGTG 807
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 808 -----GGATGTTAT 816
Db 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProGluAsnPheGlnCysAsn 278
QY 817 GGAACACTGGGATGAGTCTGTGTGATCGCGGATCTCTCAATAACACAGCATCTGTG 876
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer--- 297
QY 877 CTGAGTGTGACTGACCACACAGCGGACAGACAGACAGTGGAAACCCAAAGCCAGCTG 936
Db 298 -----ThrPheSerAspGlyArgTyrTrpThrProGlnGlnSerArgLeu 311
QY 937 AAAAAACCTGGACCGCTTGGTGTGCTTTGCCACTGATGAATACCACTGGTTACAATA 956
Db 312 HisGlyAspAspGlyTrpThrProAsnLeuAspSerAsnLysGlyLysGluIleGlnVal 331
QY 997 GATTTCGAATAAGAAAGAAATATACAGGCATTATACCACTGGA-----TCCACCATG 1050
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGGACACAATTACTATGTCTGTGCTGCTACAGAAATCTGTACAGTATGATGGCAGAAA 1110
Db 352 ThrGlnLysGlyTyrTrpValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1111 TGGACTGTGTACAGAGAGCTGTGTGGAGCAAGATAATTTCAAGAAACAAAGAT 1170
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsnAsp 389
QY 1171 TATCACAGATGTGGTAAATACTTTTCCACCAATATTACAGCTTTATTATGAGTG 1230
Db 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgile 409
QY 1231 AATCCTACCAATGGCAGCAGAAATATGCAATGAGTGTGCTGATGTCAGTTT 1290
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1291 ATTCCTAAAGTGTCTCTCCAAACTTACTCAACCTCCACTCTCTCGGAACAGCAATGAC 1350
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCAAAACACTACAGCCCTCCAAAATAGCAAGGTGCGTCCCAAAATTT----- 1404
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTrp 460
QY 1405 -----ACGCAACCACTACACCTCGCAGTAGCAATGAATTTCTGTCAGACAGAA 1455
Db 461 SerProSerAlaAlaArgLeuValSerArgSerGlyTrpPheProArgAsnProGln 480
QY 1456 CAACAACACTCCAGTCTGTATACAGAAATACCTACCTCAATGTAAACCAAGAT 1515
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----GCCTGGCTCAGTCTTGTCCCTGTGCTG 1548
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1549 GTCATGCTCTACTCTCTCATTCTCATATTAGTGTGCTGGCACTGG----- 1599

Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpIleuTyrIleGlnAsp 540
QY 1600 -----AGAAACCAAGAAAGAAACTGAAGGCACC-----TATGACTTACTACTGG 1647
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1648 GACCGGCGAGTGTGGTGAAGAAAGAAATGAAGCAGATTCTCTCCGCAAAAGCAGTGGACCAT 1707
Db 559 -----AspIleArgPheAspPro----- 565
QY 1708 GAGAAACCCCACTCTGCTATATACACAGCAAGTAAATCACCTGAGTCCCAAGAGATC 1767
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1768 ACCACAGTGTGAGCGTCACTCTGCAGAGTAT-----GCTCAGCAGCTGTGTAGGAGA 1821
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1822 ATTGTTGGTACACTTCATCAAGATCTACTTTAAACCA-----GAAGAAGGA 1869
Db 604 LeuGlyProThrValLysSerGluGluThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAGAAGCAGGCTAT-----GCAGACCTAGATCTCTTACAACCTCA 1908
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerGln 643
QY 1909 CCAGGCGAGAACTTATCATCGCT-----ATGCTGAACCACTCCCA 1950
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTrpValTyrAspHisAlaLys 663
QY 1951 A-----TTACGGGGCTGAGTATGCAACCCCAATCATCATGACATGTCAGGGCACCCCA-- 2005
Db 663 strpLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 2006 -----CAACTTCAG----- 2014
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgGln 696
QY 2015 -----TTGGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAA 2058
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 2059 CTCCTCCCACTAGTGGAACTTACATCACTCTCTCTC----- 2096
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2097 -----CAGGACTGACAGCTGCTC-----CTCAGCCAGGCCAG 2130
Db 731 lnValValArgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2131 TATGATACCCGAAAGCTGGAAAGCCAGGTCCTACCTGCCCCAGCAAAATTTGGTGTACCCAG 2190
Db 751 GlySerGluTrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2191 GTCCACAGACACACAGAAAGTATCAGGACAGGAGGATGGGAATGTATGT 2247
Db 770 Ile-----ValPheGluGlyValIleGlyLysArgSerGlyGluIleSerIle 786
RESULT 5
US-08-936-135-10
; Sequence 10, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hong
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH

Db 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 535 TCCAATTTTGGAACTGAGTCAAGTACTGCGCCAGCTGGTGTCTGCTCTCTTT 594
Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
QY 595 GCTGAGATATCTGGAACAATTCCTCATGATATAGATATCCCTGCCATTGTGCATGGCT 654
Db 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 655 GGTGTGATGATGAGAGTAGTGTCAACACACGTTGGGGGCCAAATCAGT----- 702
Db 182 LysProArgMetGluIleIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu 201
QY 703 -----GTTGTAATTAGTAAGTATATCC----- 726
Db 202 GlnValGlyGluGlyAspCysLysTyrAspTyrLeuAspIleTyrAspGlyIleProHis 221
QY 727 -----TATTATGAAGTCTTGGCTAACACGTCACATCTGTG 765
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSer 241
QY 766 GTGGACACTTATCTACAAGTCTTTTACATTTAAGACAAGT----- 807
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 808 -----GGATGTTAT 816
Db 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProGluAsnPheGlnCysAsn 278
QY 817 GGAACATGGGATGGAGTGTGTGATCGCGATCTCTCAATAACAGCATCATCTGTG 876
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSer 297
QY 877 CTGGAGTGGACTGACCACAGGCGCAAGACAGTGTGGAACCCAAAGGCCAGGCTG 936
Db 298 -----ThrPheSerAspGlyArgTyrThrProGlnGlnSerArgLeu 311
QY 937 AAAAAACCTGGACCGCTTGGCTGTCTTTGCCACTGATGAATACCAAGTGTTCACAAATA 996
Db 312 HisGlyAspAspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnVal 331
QY 997 GATTGTGAATAGGAAAGAAATAACAGGCATTATACCACTGGA-----TCCACCATG 1050
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGGAGCACATTTACTATGTCTGCTGCCTACAGATCTGTACATGATGATGGCGCAAAA 1110
Db 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1111 TGGACTGTGTACAGAGAGCTGTGTGGAGCAAGATAGATATTTCAAGGAAACAAAGAT 1170
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsp 389
QY 1171 TATCACCAGGATGGCGTAATAACTTTTGGCCACCAATATTGACGTTTATTAGAGTG 1230
Db 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1231 ATTCCTACCCATGCGCAGAAATTCGCATGAAATGGAGCTCTCGGATGTCAGTTT 1290
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1291 ATTCCTAAGGTGCTGCTCCAAACTTACTCAACCTCCACCTCCCTCGGAACAGCAATGAC 1350
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCAAAACACTACAGCCCTCCAAAATAGCCAAAGTGTGCCCCCAAAATTT----- 1404
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerThrArgGluTyrLeuTrp 460
QY 1405 -----ACCAACCACTACAACTCGCAGTAGCAATGATTTCTTCGCACAGACAGAA 1455
Db 461 SerProSerAlaAlaArgLeuValSerArgSerArgGlyTyrPheProArgAsnProGln 480

QY 1456 CAACAACATGCCAGTCTCTGATATCAGAAATACTACCGTACTCCAAATGTAACCAAGAT 1515
Db 481 AlaGlnProGlyGlyGluIleTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----GCGCTGGCTGCAGTTCTTGTCTCCCTGTGCTG 1548
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1549 GTCATGGTCTCTACTACTCTCTCATTTTCATATTAGTGTGCTGTGGCACTGG----- 1599
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTyrGluTyrIleGlnAsp 540
QY 1600 -----AGAAACAGAAAGAAAACCTGAAGCCACC-----TATGACTTACTCTTACTGG 1647
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1648 GACCGGGCAGTTGGTGGAAAGGAATGAACAGTCTTCTCTGCAAAAGCAGTGGACCAT 1707
Db 559 -----AspIleArgArgPheAspPro----- 565
QY 1708 GAGAAACCCCACTTCTGCTATACAGCAGCAAGATTAACTACCTGAGTCCCAAGAGAAGTC 1767
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTyrSerProAlaGlyIle 583
QY 1768 ACCACAGTCTGCGAGGCTGCTGCGAGAGTAT-----GCTCAGCCACTGGTAGGAGGA 1821
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTyrThrAspSerLysProThrValGluThr 603
QY 1822 ATTGTTGGTACATTCATCAAGATCTACCTTTAAACCA-----GAAGAAGGA 1869
Db 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAAGAAGCAGGCTAT-----GCAGACCTAGATCTCTTCAACTCA 1908
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerG1 643
QY 1909 CAGGGCAGGAGTTTATCATGCT-----ATGTGAACCACTCCCA 1950
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTyrValTyrAspHisAlaLys 663
QY 1951 A---TTACGGGGCTGAGTATGCAACCCCAATCATCATGAGACATGTCAGGCAACCCCA-- 2005
Db 663 sTrpLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 2006 -----CAACTTCAG----- 2014
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgG1 696
QY 2015 -----TTGTCACCCCTCCACATCCACTTTCAGGCTACGGGGAACCAA 2058
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 2059 CTTCCCCCACTAGTGGGAACCTTACAATACACTCTCTCTC----- 2096
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2097 -----CAGGACTGACAGCTGCTC-----CTCAGCCCAAGCCGAG 2130
Db 731 lnValValArgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2131 TATGATACCCCGAAGCTGGGAAGCCAGGCTACTGCTGCCCCAGCAAGTGTGTACCCAG 2190
Db 751 GlySerGlnTyrLysHisGlyArgIleLeuLeuProSerTyrAsp---MetLutyrGln 769
QY 2191 GTGCACAGACACACAAGAAGTATCAGGAGCAGGAAGGATGGGAA 2238
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu 783
RESULT 7
; US-08-936-135-14
; Sequence 14, Application US/08936135
; Patent No. 6054293

QY 364 GAAATAGCAATACACTGCTGGCTGGGTTGCCAAATGAAC-----CATTCAATTGAA 414
Db 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProThrIleIle 114
QY 415 TCAAAAGCAATGAATACATTCGCTGTCATGAGTGGAATCCATGCTTCTGGACGGGA 474
Db 115 SerSerGlySerValLeuTyriIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly 134
QY 475 TTTTGGCCTCAVACTCTGTATAGATAACAAGACTAATTACTGTTTGGACACTGCA 534
Db 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 535 TCCAAATTTTGGAACTAGTACTGAGTAACTGCCAGCTGGTGTCTGCTTCCCTTT 594
Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
QY 595 GGTGAGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTCGCCATGTCATGGCT 654
Db 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 655 GGTGTGCATGCAGAGTAGTGTCAACACAGTTGGGGCGGCCAAATCACT----- 702
Db 182 LysProArgMetGluIleIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu 201
QY 703 -----GTTCTAATTACTAAAGTATTCCC--- 726
Db 202 GlnValGlyGluCysLysTyrAspTyrPheLeuAspIleThrPheGlyIleProHis 221
QY 727 -----TATTGAAAGTCTTTGGCTAACACAGCTCACATCTGTG 765
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSerSer 241
QY 766 GTGGGACACTTACTACAGTCTTTTACATTTAAGCAAGT----- 807
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 808 -----GGATGTTAT 816
Db 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProProGluAsnPheGlnCysAsn 278
QY 817 GGAACACTGGGATGAGTGTGTGATCGCGGATCCTCAATAACAGCATCATCTGTG 876
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer--- 297
QY 877 CTGAGTGGACTGACACACAGGCGGAGACAGACAGTTGGAAACCCAAAGCCAGGCTG 936
Db 298 -----ThrPheSerAspGlyArgTyrThrProGlnGlnSerArgLeu 311
QY 937 AAAAAACCTGGACCGCTTGGCTGCTTTTGGCCACTGATGAATACCAAGTGGTTACAATA 996
Db 312 HisGlyAspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnVal 331
QY 997 GATTGTAATGAAGAAAGAAATAACAGGCAATTATACCACTGGA-----TCCACCATG 1050
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGGACACAAATFACATGCTGCTGCTACAGAAATCCTGTACAGTATGATGGCAGAAA 1110
Db 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1111 TGGACTGTGACAGAGCCCTGCTGGACCAAGATTAAGTATTTCAGGAACAACAAGAT 1170
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsnAsp 389
QY 1171 TATCACAGGATGCTGGTGAATAACTTTTGGCCACCAATTATGACCGTTTATTAGAGTG 1230
Db 390 AlaThrGluValValLeuAsnLysLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1231 AATCCTACCAAGGACGACGAAATGTCATGAAATGGAGTGTGCTGGATGTCACTTT 1290
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1291 ATTCCTAAAGGTGCTCTCCAAACTACTCACTCCACTCCACTCTCGAACAGCAATGAC 1350

Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCAAAACACTACAGCCCTCCAAATAAGCCAAAGGTGTCGCCCAAAATTT----- 1404
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTrp 460
QY 1405 -----ACGCAACCACTACAACTCGACCTGAGCAATGAATTCCTGCACAGACAGAA 1455
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTyrPheProArgAsnProGln 480
QY 1456 CAAACAACTCCAGTCTGTATATCAGAAATACCTACCTAACTCAATGTAACCAAGAT 1515
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----GCCTGGCTGCAGTCTCTGTGCTGCTGGCACTGG----- 1599
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1549 GTCATGGTCTCTACTCTCATCTCATATTAATAGTGTGCTGCTGGCACTGG----- 1599
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTyrGluTyrIleGlnAsp 540
QY 1600 -----AGAACAGAAAGAAACTGAAGGCACC-----TATGACTTACCTTACTGG 1647
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMethHisTyrAspThrPro----- 558
QY 1648 GACCGGCGAGGTGGTGGAAAGAAATGAAGCAGTTTCTCTGCAAAAGCAGTGGACCAT 1707
Db 559 -----AspIleArgArgPheAspPro----- 565
QY 1708 GAGAAACCCAGTTCCTATAGCAGCAGGAGAGTAAATCACTGATGTCACAAAGAGATC 1767
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTyrSerProAlaGlyIle 583
QY 1768 ACCACAGTCTGAGGCTGACTCTGCAGAGTAT-----GCTCAGCCACTGTAGGAGGA 1821
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTyrThrAspSerLysProThrValGluThr 603
QY 1822 ATTGTTGGTACACTTCATCAAGATCTACCTTTAAACCA-----GAAGAGGA 1869
Db 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAGAAGCAGGCTAT-----GCACACCTAGATCTCTCAACTCA 1908
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerG1 643
QY 1909 CCAGGGCAGCAAGTTTATCATGCT-----ATGCTGAACCACTCCCA 1950
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTyrPvalTyrAspHisAlaLy 663
QY 1951 A---TTACGGGGCTGAGTATGCAACCCCAATCATCATGACATGTGAGGCACCCCA-- 2005
Db 663 sTrpLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 2006 -----CAACTTCAG----- 2014
Db 676 nAspArgThrPheProAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArgG1 696
QY 2015 -----TTGGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACAA 2058
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 2059 CTTCCCCCACTAGTGGGAACCTTCAATACACTTCTCTCT----- 2096
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2097 -----CAGACTGACAGTGTCTC-----CTCAGCCCGAGGCCAG 2130
Db 731 InValValArgGluAlaSerGlnGlnSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2131 TATGATACCCGGAAGCTGGGAAGCCAGGTCTACCTGCCCGCCAGCAATTTGGTGTACCAG 2190

Qy	1180	GATGTGCGTATAACTTTTGGCCACCAATTATGTGACGCTTTATTATTAGATGGATCTCTACC	1233
Db	393	ValValLeuAsnLysLeuHisAlaProLeuLeuThrArgPheValArgLeuArgProGln	412
Qy	1240	CAATGGCAGCAGAAAATGCCATGAAATGGAGCTCTCGGATGTCAGTTTATTCCTAAA	1299
Db	413	ThrTrpHisSerGlyIleAlaLeuArgLeuLeuPheGlyCys	427
Qy	1300	GGTCGTCTCCAAAATTTACTCACTCCACCTCTCGGAACAGCAATGACCTCAAAAAC	1359
Db	428	-----ArgValThrAspAlaProCysSerAsnMetLeuGlyMetLeuSerGly	443
Qy	1360	ACTCAGCCCTCCAAAATGACCAAGGTCGTGCCCAAAATTT	1404
Db	444	LeuIleAlaAspSerGlnIleSerAlaSerSerThrGlnGluTyrrLeuTrpSerProSer	463
Qy	1405	AGCAACCACTACAACCTCGCAGTAGCAATGATTTCTTCGACAGCAGACAACAACACT	1464
Db	464	AlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgIleProGlnAlaGlnPro	483
Qy	1465	GCCAGTCCTGATATCAGAAATACTACCGTAACGCCAAATGTAACCAAGATGTA	1518
Db	484	GlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGlyValIlelle	503
Qy	1519	-----GCGTCGGTGCAGATCTTCTTCCTCCTGCTGGTCGATGGTC	1557
Db	504	GlnGlyAlaArgGlyClyAspSerIleThrAlaValGluAlaArgAlaPheValArgLys	523
Qy	1558	CTCACTACTCTCATCTCATATTAGTGTGTGGCTGGCACTGG	1602
Db	524	PheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTrpIleGlnAspProArgThr	543
Qy	1603	AAACAGAAAGAAAACACTGAAGGCACC	1656
Db	544	GlnGlnProLysLeuPheGluGlyAsnMetHisTyrAspThrPro	558
Qy	1657	GTTGGTGGAAAGGAATGAAGCAGTTTCTTCCTGCAAAAGCAGTGGACCATGAGGAAC	1716
Db	559	-----AspIleArgArgPheAspPro	566
Qy	1717	CCAGTTCGTATAGCAGCAGCAGGAAGTTAATCACTGTGATCCAAAGAGTCAACCACAGT	1776
Db	567	ProAlaGlnTrpValArgValTyrrProGluArgTrpSerProAlaGlyIleGlyMetArg	586
Qy	1777	CTGCAGGCTGACTCTGCAGAGTAT	1830
Db	587	LeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValLysThrLeuGlyPro	606
Qy	1831	ACACTTCATCAAGATCTACCTTTAAACCA	1878
Db	607	ThrValLysSerGluGluThrThrThrProTyrProThrGluGluGluAlaThrGluCys	626
Qy	1879	GGC-----TATGCAGACCTAGATCTTACACTCACCAGGGCAGGAAG	1921
Db	627	GlyGluAsnCysSerPheGluAspAspLysAsp	646
Qy	1922	-----TTTATATGCGCT	1956
Db	646	AsnPheAspPheLeuGluGluProCysGlyTrpMetTyrrAspHisAlaLysTrpLeuAr	666
Qy	1957	GGCCCTGAGTATGCAACCCCAATCATCATGGACATGTCAGGGCACCACCACA	2008
Db	666	g-----ThrThrTrpAlaSerSerSerProAsnAspArgTh	679
Qy	2008	-----	2008
Db	679	rPheProAspAspArgAsnPheLeuArgLeuGlnSerAspSerGlnArgGluGlyGlnTy	699
Qy	2009	-----CTTCAGTTGGTCAGCCCTCCACATCCACTTTCAAGGTCAGGGGAACCACTCCCCCA	2067
Db	699	rAlaArgLeuIleSerProProValHisLeu	714

QY	2068	CTAGTGGGAACCTTCAACATACACTTCTCTC-		---	2098
Db	714	alCysMetGluPheGlnAlaThrGlyArgGlyValalaLeuGlnValVala	734		
QY	2097	-----CAGACTGCACAGCTGCTC-		---	2109
Db	734	rgGluLaserGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlyGlyGlu	753		
QY	2140	CGAAAGCTGGGAAGCCAGGTTACTCTCCCGACAGCAATTTGGTGTTACCGAGTGCCACAG	2199		
Db	754	TpLysHisGlyArgIleIleLeuProSerTyraSp---MetGluTyrGlnIle----	770		
QY	2200	AGCACACAAGAGTATCAGGACGACGAAGGATGGGAA	2238		
Db	771	ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu	783		
RESULT 11					
US-08-936-135-6					
; Sequence 6, Application US/08936135					
; Patent No. 6054293					
; GENERAL INFORMATION:					
; APPLICANT: Tessier-Lavigne, Marc					
; APPLICANT: He, Zhigang					
; APPLICANT: Chen, Hang					
; TITLE OF INVENTION: Semaphorin Receptors					
; NUMBER OF SEQUENCES: 26					
; CORRESPONDENCE ADDRESS:					
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP					
; STREET: 75 DENISE DRIVE					
; CITY: HILLSBOROUGH					
; STATE: CALIFORNIA					
; COUNTRY: USA					
; ZIP: 94010					
; COMPUTER READABLE FORM:					
; MEDIUM TYPE: Floppy disk					
; COMPUTER: IBM PC compatible					
; OPERATING SYSTEM: PC-DOS/MS-DOS					
; SOFTWARE: Patent In Release #1.0, Version #1.30					
; CURRENT APPLICATION DATA:					
; APPLICATION NUMBER: US/08/936.135					
; FILING DATE:					
; CLASSIFICATION: 435					
; ATTORNEY/AGENT INFORMATION:					
; NAME: OSMAN, RICHARD A					
; REGISTRATION NUMBER: 36,627					
; REFERENCE/DOCKET NUMBER: UC97-288-2					
; TELECOMMUNICATION INFORMATION:					
; TELEPHONE: (650) 343-4341					
; TELEFAX: (650) 343-4342					
; INFORMATION FOR SEQ ID NO: 6:					
; SEQUENCE CHARACTERISTICS:					
; LENGTH: 923 amino acids					
; TYPE: amino acid					
; STRANDEDNESS: single					
; TOPOLOGY: linear					
; MOLECULE TYPE: peptide					
US-08-936-135-6					
Alignment Scores:					
Pred. No.: 9,29e-28 Length: 923					
Score: 392.50 Matches: 132					
Percent Similarity: 39.23% Conservative: 61					
Best Local Similarity: 26.83% Mismatches: 178					
Query Match: 9.47% Indels: 121					
DB: 3 Gaps: 19					
US-10-060-830-1 (1-2280) x US-08-936-135-6 (1-923)					
QY	31	CTCTCCGGTCCCTCCCTCCCTCCCACTCCTCCTCTCTCCATCCTCTGTCTC	90		
Db	1	MetGluArgGlyLeuPro-----	6		
QY	91	CTGCTCTTACTTGTCTGTCTGCTGCTGCTGAGGACGCTGGAGCCCAAGGATGATGGA	150		

Db 272 ArgGly-----AsnValAspAsnAsnThr----- 279
Qy 712 AGTAAAGGTATTCCTATTATGAAGTTCTTGGCTAAACAACGTACATCTGTGGGGA 771
Db 280 -----ProTyr-----AlaAsnSerPheThrProPheIleLys 290
Qy 772 -----CACTTATCTACAAGT 786
Db 291 AlaGlnTyrValArgLeuTyrProGlnIleCysArgArgHisCysThrLeuArgMetGlu 310
Qy 787 CTTTTCATATTAAAGCAAGTGGATTATGGAACACTGGGATGGAGTCTGTGTGATC 846
Db 311 LeuLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMetLysSerGlyHisIle 330
Qy 847 GCGATCCTCAATAACAGCATCATCTGTCTGTGGAGTGGAGTACACACAGGCAAGAG 906
Db 331 GlnAspTyrGlnIleThrAlaSerSerValPheArgThrLeuAsnMet-----AspMet 348
Qy 907 AACAGTTGGAACCCAAAGAGCCAGGCTGAAGAAACCTGGACCG-----CCTTGGGCT 960
Db 349 PheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysValAsnAlaTrpThr 368
Qy 961 GCTTTTCCACTGATGATCACCAGTGTTCACAAATAGATTGTAATAAGAGAAAGAAATA 1020
Db 369 SerGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuLeuValProThrLysVal 388
Qy 1021 ACAGCATATTAAACACTGGATCCACCATGGTGAGACCAATTAATCTATGTCTGCTAC 1080
Db 389 ThrGlyIleThrGlnGlyAlaLysAspPheGlyHisValGlnPheValGlySerTyr 408
Qy 1081 AGATCTGTACATGATGATGGCAGAAATGGACTGTGTACAGAGAGCCCTGGTGTGGAG 1140
Db 409 LysLeuAlaTyrSerAsnAspGlyGluHisTrpMetValHisGlnAspGluLysGlnArg 428
Qy 1141 CAAGATAAGATATTCAAGGAAACAAAGATTATCACCAGGATGTCGTAATAACTTTTG 1200
Db 429 LysAspLysValPheGlnGlyAsnPheAspAsnAspThrHisArgLysAsnValIleAsp 448
Qy 1201 CCACCAATTATGACGTTTATTAGATGAATCCTACCCCAATGGCAGCAGAAATGCC 1260
Db 449 ProPheIleTyrAlaArgPheIleArgIleLeuProTrpSerTrpTyrGlyArgIleThr 468
Qy 1261 ATGAAATGGAGCTGCTCGGATGT 1284
Db 469 LeuArgSerGluLeuLeuGlyCys 476

RESULT 14
US-08-659-235C-10
Sequence 10. Application US/08659235C
Patent No. 5877281
GENERAL INFORMATION:
APPLICANT: Quertemous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/659,235C
FILING DATE: 05-JUN-1996

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0034-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-659-235C-10
Alignment Scores:
Pred. No.: 2,96e-25 Length: 480
Score: 364.00 Matches: 138
Percent Similarity: 37.12% Conservative: 58
Best Local Similarity: 26.14% Mismatches: 170
Query Match: 8.79% Indels: 162
DB: 2 Gaps: 21
US-10-060-830-1 (1-2280) x US-08-659-235C-10 (1-480)
Qy 13 GCTGGCGCGCTCCCTCCCTCTCCCGCTCCCTCCCT----- 48
Db 7 AlaTrpLeuLeuValGlyLeuSerLeuGlyValProGlnPheGlyLysGlyAspIleCys 26
Qy 49 -----CCCTCTCCAACTCCTCCTCTCTCCATCCCTCTCTCTCTCTCTCTCTCTCTCTTA 99
Db 27 AsnProAsnProCysGluAsnGlyCly----- 35
Qy 100 CTTGCTCTGCTCTCTGCTCTGAGACGCTGGAGCCAGCAAGTGATGGATGGAGAC 159
Db 36 IleCysLeuSerGlyLeuAlaAspSerPheSerCysGlu-----CysProGlu 52
Qy 160 ACTGTACTAGCCCTGAGAGTGGAACCTTACATCCATAAACCACACACACCTATCCC 219
Db 53 GlyPheAlaGlyProAsnCysSerSerValValGluValAlaSerAspGluGluLysPro 72
Qy 220 -----AACAGCACTGTGTGTGAATGGAG 243
Db 73 ThrSerAlaGlyProCysIleProAsnProCysHisAsnGlyGlyThrCysGlu----- 90
Qy 244 ATCCGTGTAAGATGGGAGAGAGATTCGCATCAAAATTTGGTGACTTTGCATTTGAAGAT 303
Db 91 -----IleSerGluAlaTyrArg-----GlyAspThrPheIleGlyTyr 103
Qy 304 TCTGATCTTGT-----CACTTTAATTACTTGAGAATTTAATGGAATGGAGTCAGC 357
Db 104 ValCysLysCysProArgGlyPheAsnGlyIleHisCysGlnHisAsnIleAsnGluCys 123
Qy 358 AGAACTGAAATAGCAAAATACTGTGTCTG-----GGTTGCAAAATGAACCATCAATT 411
Db 124 GluAlaGluProCysArgAsnGlyGlyIleCysThrAspLeuValAlaAsnTyrSerCys 143
Qy 412 GAATCAAAAGCAATGAAATCACATTCGTGTTCATGAGT----- 450
Db 144 GluCysProGlyGlu-----PheMetGlyArgAsnCysGlnTyrLysCys 158
Qy 451 -----GGAATCCAT-----GTTCTGACGCGGATTTTGGCCCTCA--- 486
Db 159 SerGlyHisLeuGlyIleGluGlyIleIleSerAsnGlnGlnIleThrAlaSer 178
Qy 487 -----TACTCTCTTATAGATAAA 504
Db 179 AsnHisArgAlaLeuPheGlyLeuGlnLysTrpTyrProTyrTyrAlaArgLeuAsnLys 198
Qy 505 CAAGATCTAATTAATCTGTTTGGACACTGCATCCAAT----- 540
::: ::::: ||| |||

Db 199 LysGlyLeuIleAsnAlaTrpThrAlaAaGluAsnAspArgTrpProTrpIleGlnIle 218
QY 541 -----TTTTT 546
Db 219 AsnLeuGlnArgLysMetArgValThrGlyValIleThrGlnGlyAlaLysArgIleGly 238
QY 547 GAACCTGAGTTCAGTAAG-----TACTGCCACGCTGTTGTCTGCTTCCT 591
Db 239 SerProGluTyrIleLysSerTyrLysIleAlaTyrSerAsnAspGlyLysThrTrpAla 258
QY 592 TTTGCTGAGATCTCTGGAACAATTCCTCATGATAGAGATTCCTCGCCATTCGTCATG 651
Db 259 MetTyrLysValLysGlyThr-----AsnGluGluMetValPhe 271
QY 652 GCTGGTGCATGCAGGAGTAGTCTCAACACAGTTCGGGGGCCCAATCACTGTGTAATT 711
Db 272 ArgGly-----AsnValAspAsnAsnThr----- 279
QY 712 AGTAAGGTATTCCTATTATGAAGTCTTTGGCTAAACAACGTCACATCTGTGGGA 771
Db 280 -----ProTyr-----AlaAsnSerPheThrProPheLys 290
QY 772 -----CACTTATCTACAAGT 786
Db 291 AlaGlnTyrValArgLeuTyrProGlnIleCysArgHisCysThrLeuArgMetGlu 310
QY 787 CTTTTFACATTAAGACAAGTGGATCTTATGGAACACTGGGATGGAGTCTGTGTGATC 846
Db 311 LeuLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMetLysSerGlyHisIle 330
QY 847 GCGGATCCTCAATACACATCATCTGTGCTGGAGTGGACTACACACAGGCGCAAGAG 906
Db 331 GlnAspTyrGlnIleThrAlaSerSerValPheArgThrLeuAsnMet-----AspMet 348
QY 907 AACAGTTGGAACCCCAAAAGCCAGCTGCAAAAACCTGGACCG-----CCTTGGCT 960
Db 349 PheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysValAsnAlaTrpThr 368
QY 961 GCTTTTGCCACTGATGAATACCACTAGTGGTGTACAAATAGATTGAATAAGGAAAAAATA 1020
Db 369 SerGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuLeuValProThrLysVal 388
QY 1021 ACAGGATTAATACCACTGGATCCACCATGGTGGGACACAAATFACATGTGTGCTGCTAC 1080
Db 389 ThrGlyIleThrGlnGlyAlaLysAspPheGlyHisValGlnPheValGlySerTyr 408
QY 1081 AGAATCCTGACAGTATGATGGCAGAAATGACTGTGTACAGAGAGCCGTGTGGAG 1140
Db 409 LysLeuAlaTyrSerAsnAspGlyGluHisTrpMetValHisGlnAspGluLysGlnArg 428
QY 1141 CAAGATAAGATATTCAAGAAACAAGATTATCACCAGGATGCGGTAATAACTTTTG 1200
Db 429 LysAspLysValPheGlnGlyAsnPheAspAsnAspThrHisArgLysAsnValIleAsp 448
QY 1201 CCACCAATTAATCCAGCTTTTATAGATGAATTCCTACCAATGGCAGCAGAAATGGCC 1260
Db 449 ProProIleTyrAlaArgPheIleArgIleLeuProTrpSerTyrGlyArgIleThr 468
QY 1261 ATCAAAATGGAGCTCGCATCT 1284
Db 469 LeuArgSerGluLeuLeuGlyCys 476

RESULT 15
US-08-480-229C-14
; Sequence 14, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quermous, Thomas
; APPLICANT: Hogan, Bridg
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; CELL LOCUS-1

NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,229C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-229C-14

Alignment Scores:
Pred. No.: 1,87e-24 Length: 513
Score: 355.50 Matches: 137
Percent Similarity: 38.45% Conservative: 51
Best Local Similarity: 28.02% Mismatches: 180
Query Match: 8.58% Indels: 121
DB: 2 Gaps: 20

US-10-060-830-1 (1-2280) x US-08-480-229C-14 (1-513)

QY 68 CCTCTCCATGC-----CTCTGT-----TCCTCCTGCTCTTACTTCTCC 106
Db 60 ProAsnProCysGluAsnGlyGlyIleCysLeuProGlyLeuAlaValGlySerPheSer 79
QY 107 TGCTCCTGC-----TGCTCGAGGACGCTGGAGCCCAAG 142
Db 80 CysGluCysProAspGlyPheThrAspProAsnCysSerValValGluValAla--S 99
QY 143 GTGATGATGTGCACACACTGTACTAGCCCTGAGAGTGGAAACCTTACATCCATAACT 202
Db 99 erAspGluGluProThrSerAlaGlyProCysThr----- 111
QY 203 ACCCAGACAGCTATCCACACAGCACTGTTGTGAATGGGAGATCCGCTGTAAGATGGGAG 262
Db 112 --ProAsnProCysHisAsnGlyGlyThrCysGlu-----IleSerG 125
QY 263 AGAGAGTTCGCATCAAAATTTGGTACATTTGACATTTGAAGATTTCTGATCTTGT-----C 316
Db 125 LuAlaTyrArg-----GlyAspThrPheIleGlyTyrValCysLysCysProArg 142
QY 317 ACTTTAATTAATTCAGAAATTTAATGAATTTGAGTTCACAGACACTGAATAGGCAAT 376
Db 142 lyPheAsnGlyIleHisCysGlnHisAsnIleAsnGluCysGluValGluProCysLys 162
QY 377 ACTGTGCTCG-----GGGTTGCAAAATGAACCATTTCAATTAAGGCAATGAAA 430
Db 162 snGlyGlyIleCysThrAspLeuValAlaAsnTyrSerCysGluCysProGlyGlu---- 180
QY 431 TCACATTCGCTGTTTCATGAGT-----GGATTC 457

Job time : 46.2425 secs

```
Db 181 -----PheMetGlyArgAsnCysGlnTyrLysCysSerGlyProLeuGlyIleG 197
|||||
QY 458 AT-----GTTTCTGGACGCGGATTTTGGCCTCA----- 486
|||||
Db 197 LuGlyGlyIleIleSerAsnGlnGlnIleThrAlaSerThrHisArgAlaLeuPheG 217
|||||
QY 487 -----TACTCTGTTATAGATAACAAGATCTAATTAAGTCTTGT 523
|||||
Db 217 lYLeuGlnLysTrpTyrProTyrTyrAlaArgLeuAsnLysGlyLeuIleAsnAlaT 237
|||||
QY 524 TGCACACTGCATCCCAATTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGGTGTGC 583
|||||
Db 237 rPThrAlaAlaGluAsn-----AspArgTrpLysArgTrpIleGlnIleAsn--L 253
|||||
QY 584 TGCTTCCTTTTGGTGTAGATATCTGGAACAATTCCTCATGATATAGA----- 630
|||||
Db 253 euGlnArgLysMetArgValThrGlyValIleThrGlnGlyAlaLysArgIleGlySerP 273
|||||
QY 631 -----GATTCTCGCCATTGTCATGG 652
|||||
Db 273 roGluTyrIleLysPheTyrLysIleAlaTyrSerAsnAspGlyLysThrTrpAlaMetT 293
|||||
QY 653 CTGGTGTGCATGCAGGAGTAGTCTCAACACAGTGTGGCGGCCAAATCAGTGTGTAATTA 712
|||||
Db 293 yrLysValLysGlyThrAsnGluAspMetValPheArgGlyAsn-----IleA 309
|||||
QY 713 GTAAGGTATTCCTATTATGAAAGTCTTGGCTAACACAGTCACATCTGTGGTGGGA- 771
|||||
Db 309 sPAsnAsnThrProTyr-----AlaAsnSerPheThrProIleLysA 324
|||||
QY 772 -----CACTTATCTACAAGTC 787
|||||
Db 324 laGlnTyrValArgLeuTyrProGlnValCysArgArgHisCysThrLeuArgMetGluL 344
|||||
QY 788 TTTTACATTAAAGACAGTGGATGTATGGAACACTGGGGATGGAGTCTGGTGTGATCG 847
|||||
Db 344 euLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMetLysSerGlyHisIleG 364
|||||
QY 848 CGATCTCTCAATAACAGCATCATCTGCTGGTGGAGTGGAGTACCACACAGGGCAAGAGA 907
|||||
Db 364 lNAspTyrGlnIleThrAlaSerSerIlePheArgThrLeuAsnMet-----AspMetP 382
|||||
QY 908 ACAGTTGGAAACCCAAAGCCAGGCTGAAACCTGGACCG-----CCTTGGGCTG 961
|||||
Db 382 heThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysValAsnAlaTrpThrS 402
|||||
QY 962 CTTTGGCACTGATGAATACAGTGGTTACAATAGATTTGAATAAGGAAAGAAATAA 1021
|||||
Db 402 erGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuLeuValProThrLysValT 422
|||||
QY 1022 CAGCATTAATACACCTGGATCCACCATGCTGGAGCACAATTAATCTGTCTGCTGTACA 1081
|||||
Db 422 hrGlyIleIleThrGlnGlyAlaLysAspPheGlyHisValGlnPheValGlySerTyrL 442
|||||
QY 1082 GAATCTCTACATGATGGCGCAGAAATGGACTGTGTACACAGAGCCCTGGTGTGGAGC 1141
|||||
Db 442 ysLeuAlaTyrSerAsnAspGlyGluHisTrpThrValTyrGlnAspGluLysGlnArgL 462
|||||
QY 1142 AAGATAAGATATTTCAGGAAACAAAGATTATCACCAGGATGTCGGTAATAACTTTTTCG 1201
|||||
Db 462 ysAspLysValPheGlnGlyAsnPheAspAsnAspThrHisArgLysAsnValIleAspP 482
|||||
QY 1202 CACCAATATTGACAGTGTATTTATTAGATGAATCCTACCCCAATGGCAGCAAAATGCCA 1261
|||||
Db 482 roProIleTyrAlaArgHisIleArgIleLeuProTrpSerTrpTyrGlyArgIleThrL 502
|||||
QY 1262 TGAATAATGGAGCTGCTCGGATGT 1284
|||||
Db 502 euAlaSerGluLeuGlyCys 509
|||||
```

Search completed: January 21, 2003, 09:50:50

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model
Run on: January 21, 2003, 02:16:00 ; Search time 54.9883 Seconds
(without alignments)
10613.845 Million cell updates/sec

Title: US-10-060-830-2
Perfect score: 3969
Sequence: 1 atgcctctgtctctctgct.....ttttaagaataccttga 2190

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 04
Maximum Match 1008
Listing first 45 summaries

Command line parameters:
-MODEL=frame+.n2p.model -DEV=xlh
-O=/cnp2.1/USPTO.spool/US10060830/runat.16012003.091238.26833/app_query.fasta.1.6933
-DB-A.Geneseq.101002-QFMT-fastan-SUFFIX=n2p.rag-MINMATCH=0.1-LOOPCL=0
-LOOPEXT=0-UNITS=bits-START=1-END=1-MATRIX=BLOSUM62-TRANS-human40.cdi
-LIST=45-DLOCALIGN=200-THR_SCORE=pct-THR_MAX=100-THR_MIN=0-ALIGN=15
-MODE-LOCAL-OUTFMT=ptc-NORM-ext-HEAPSIZE=500-MINLEN=0-MAXLEN=2000000000
-USER=US10060830.ecgn.1.1.230@runat.16012003.091238.26833-NCPU=3
-NO_XLPXY-NO_MMAPP-LARGEQUERY-NEG_SCORES=0-WAIT-LONGLOG-DEV_TIMEOUT=120
-WARN_TIMEOUT=30-THREADS=1-XGAPOP=10-XGAPEXT=0.5-FGAPOP=6-FGAPEXT=7
-XGAPOP=10-YGAPEXT=0.5-DELOP=6-DELEXT=7

Database : A_Geneseq.101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	3572	90.0	669	21	AA70539 Human Factor 8 Hom
2	2806	70.7	583	22	AA70539 Human colon cancer
3	1841.5	46.4	385	23	AAE22716 Human neuropilin-H
4	1841.5	46.4	385	23	AAU79460 Human neuropilin-H
5	1841.5	46.4	385	23	ABE97386 Novel human protei
6	1750.5	44.1	365	23	AAE22721 Human neuropilin-H
7	1042	26.3	715	22	AAU00670 Human TANGO 229 po
8	1039	26.2	197	22	ABG01298 Novel human diagno
9	924.5	23.3	539	22	AAU00630 Novel human protei
10	924.5	23.3	586	22	AAU00629 Novel human protei
11	860.5	21.7	487	22	AAU00628 Polypeptide isolat
12	574.5	14.5	503	21	AB19126 Human neuropilin-H
13	540.5	13.6	398	23	AAE22715 Human neuropilin-H
14	540.5	13.6	398	23	AAU79459 Human neuropilin-H
15	421.5	10.6	925	20	AAW96308 Neuropilin-2. Rat
16	414.5	10.4	901	20	AAW96256 Mouse semaphorin r
17	414.5	10.4	906	20	AAW96257 Mouse semaphorin r
18	414.5	10.4	909	20	AAW96249 Rat semaphorin rec
19	412.5	10.4	909	20	AAW96250 Mouse semaphorin r
20	412.5	10.4	909	21	AAE24214 Mouse soluble neur
21	412.5	10.4	914	20	AAW96251 Mouse semaphorin r
22	412.5	10.4	926	20	AAW96252 Mouse semaphorin r
23	412.5	10.4	931	20	AAW96253 Mouse semaphorin r
24	412	10.4	75	22	ABE2217 Peptide #9723 enco
25	412	10.4	75	22	AAW63102 Human brain expres
26	412	10.4	75	22	AAW75913 Human bone marrow
27	412	10.4	75	22	AAW36024 Peptide #10061 enc
28	412	10.4	75	23	ABG45349 Human peptide enco
29	408.5	10.3	889	21	AAE24216 Soluble neuropilli
30	404.5	10.2	957	22	AAW25770 Human protein sequ
31	400.5	10.1	538	22	AAU02950 Angiotensin conver
32	400.5	10.1	600	22	AAU02948 Angiotensin conver
33	400.5	10.1	644	20	AAW06319 Human soluble neur
34	400.5	10.1	840	22	AAU02949 Angiotensin conver
35	400.5	10.1	856	22	AAE62478 Human NP-1 recepto
36	400.5	10.1	923	20	AAW06317 Human neuropilin-1
37	400.5	10.1	923	20	AAW23247 Human VEGF165R/NP
38	400.5	10.1	923	20	AAW96246 Human semaphorin r
39	400.5	10.1	923	23	AAE22717 Human neuropilin-1
40	400.5	10.1	924	22	AAE62476 Human semaphorin r
41	398.5	10.0	909	20	AAW96254 Mouse semaphorin r
42	398.5	10.0	923	20	AAW14562 Human VEGF(165)R/N
43	398.5	10.0	926	20	AAW96255 Human semaphorin r
44	398.5	10.0	926	23	AAE22718 Human neuropilin-2
45	398.5	10.0	931	20	AAW06318 Human neuropilin-2

ALIGNMENTS

RESULT 1
AA70539
ID AAY70539 standard; Protein; 669 AA.
XX
AC AAY70539;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human Factor 8 Homologue.
XX
KW Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW cerebroprotective; therapeutic; coagulation related disorder;
KW haemophilia; stroke; screening.
XX
OS Homo sapiens.
XX
PN WO200012532-A1.
XX
PD 09-MAR-2000.

XX 20-AUG-1999; 99WO-US19047.
 XX 31-AUG-1998; 98US-0098521.
 XX (ELIL) LILLY & CO ELI.
 XX Rostek PRJ, Su W, Li XM;
 XX WPI: 2000-256580/22.
 XX N-PSDB; AAZ51872.
 XX Factor 8 homolog polypeptides and nucleic acids encoding them for
 XX treating coagulation related disorders such as hemophilia and stroke
 XX Claim 3; Page 64-66; 68pp; English.
 CC The present sequence is a human Factor 8 homologue (F8H),
 CC a coagulation cofactor which is selectively expressed in
 CC haematopoietic, heart and reproductive tissues. It has haemostatic and
 CC cerebroprotective activities. The F8H contains a factor 5/8 signature
 CC and is useful as a therapeutic for treating coagulation related diseases
 CC such as haemophilia and stroke. The nucleic acid is useful as
 CC hybridisation probe and amplification primer for detecting deficiencies
 CC in the level of F8H mRNA, for screening F8H gene mutations and for
 CC monitoring regulation of gene expression. Fragments of the nucleic acid
 CC are also useful as diagnostic probes and primers, and can be used in
 CC screening methods such as those using DNA chips. The present sequence is
 CC also useful as a target to screen therapeutically useful modulators
 CC of the F8H.
 XX SQ Sequence 669 AA;
 Alignment Scores:
 Pred. No.: 0 Length: 669
 Score: 3572.00 Matches: 669
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 90.00% Indels: 0
 DB: 21 Gaps: 0
 US-10-060-830-2 (1-2190) x AAY70539 (1-669)
 QY 181 ATGGAGAGAGAGATCGGATCAAAATTTGGTGACTTTGACATGAAGATCTGATCTTGT 240
 Db 1 MetGlyGluArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCys 20
 QY 241 CACTTAATTAATCTGAGATTTATATGAATTCGAGTCAGACAGAACTGAATAGGCAAA 300
 Db 21 HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys 40
 QY 301 TACTGTGTCTGGGTTCGCAATGAACCATTCATTAATGAATCAAAAGCAATGAATCACA 360
 Db 41 TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr 60
 QY 361 TTGCTGTTTCATGAGTGAATCCATGTTCTTGACCGCGGATTTTGGCTCATACTCTGT 420
 Db 61 LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal 80
 QY 421 ATAGATAAACAAGATCTAATTACTTTGACACTGCGATCCCAATTTTGGAACTGAG 480
 Db 81 IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu 100
 QY 481 TTCAGTAAGTACTGCCAGCTGGTGTCTGCTTCCTTTTGGTGAATATCTGCAACAAT 540
 Db 101 PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIle 120
 QY 541 CCTCATGATATAGAGATCTCGCCATGTCATGCGTGTGTCATGCGAGGAGTAGTG 600
 Db 121 ProHisGlyTyrArgAspSerProLeuCysMetAlaGlyValHisAlaGlyValVal 140
 QY 601 TCACACCTTTGGCGGCCAATCAGTGTGTAATAGTAAGGTATCCCTATTATGAA 660
 Db 141 SerAsnThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGlu 160
 QY 661 AGTTCCTTTGGCTAACACGTCACATCTGTGTGGGACACTTATCTACAGTCTTTTACA 720
 Db 161 SerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr 180
 QY 721 TTTAAGACAGTGTGATGTTATGGAACTGCGGATGGAGTCTGGTGTGATCCGGATCCT 780
 Db 181 PheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspPro 200
 QY 781 CAAATAACAGCATCTGTCTGTGAGTGGACTGACACACAGGCAAGACAGTGTGG 840
 Db 201 GlnIleThrAlaSerSerValLeuGluThrPheAspHisThrGlyGlnGluAsnSerTrp 220
 QY 841 AAACCCAAAAGCCAGGCTGAAAACCTGGACCGCTGGGGCTGCTTTCCCACTGAT 900
 Db 221 LysProLysLysAlaArgLeuLysLysProGlyProTrpPheAlaAlaPheAlaThrAsp 240
 QY 901 GAATACCACTGGTTACAATAGATTGTAATAGGAAAAGAAATAACAGGCATTATAACC 960
 Db 241 GluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThr 260
 QY 961 ACTGGATCCACCTGGTGAGCACAAATTAATGCTGTCTGTCTACAGATCTCTGACAGT 1020
 Db 261 ThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSer 280
 QY 1021 GATGATGGCAGAAATGGACTGTGTACAGAGAGCTGTGTGGGACAGATGAATATTT 1080
 Db 281 AspAspGlyGlnLysTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePhe 300
 QY 1081 CAAGGAACAAGATTATCACCAGGATGCGTATTAATCTTTTGGCCACCAATTTATGCA 1140
 Db 301 GlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProIleIleAla 320
 QY 1141 CGTTTATTAGATGAATCCCTACCAATGGCAGACAGAAATTCCTCAATGAATGGAGCTG 1200
 Db 321 ArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeu 340
 QY 1201 CTCGATGTCAGTTTATTCCTAAAGTGTGCTCTCCAAAACCTACTCAACCTCACCTCT 1260
 Db 341 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProPro 360
 QY 1261 CGGACACCAATGACCTCAAAAACACTACAGCCCTCCAAAATAGCCAAAGTCTGTGC 1320
 Db 361 ArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla 380
 QY 1321 CCAAAATTTAGCAACCACTACAACTCCAGTAGCAATGAATTTCTCCACAGACAGAA 1380
 Db 381 ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 400
 QY 1381 CAACAACCTGCCAGTCTGATATCAGAAATACTACCGTAACCTCAAAATGAACCAAGAT 1440
 Db 401 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 420
 QY 1441 GTAGCGTGGCTGCGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 Db 421 ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 440
 QY 1501 ATATTAGTGTCTGCTGGCACTGGAGAAACAGAAAAGAAAACCTGAAGCAGCTATGAC 1560
 Db 441 IleLeuValCysAlaThrHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrAsp 460
 QY 1561 TTACCTTACTGGAGCGGCGAGTGGTGGTGAAGGAAAGTGAAGCAGTCTTCTCTGCAAAA 1620
 Db 461 LeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLys 480
 QY 1621 GCAGTGGACCATGAGAAACCCAGTTCGCTATAGCAGCAGCAGCAAGTGAATCACTGAGT 1680
 Db 481 AlaValAspHisGluGluThrProValArgTyrSerSerSerGluValAsnHisLeuSer 500
 QY 1681 CCAAGCAAGATCACCACAGTGTGTCAGGCTGACTTCGACAGATGATGCTCAGCAGCTGGTA 1740
 Db 501 ProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuVal 520

Db 245 SerSerLeuAlaAsnValThrSerValValGlyHisLeuSerThrSerLeupheThr 264
QY 721 TTTAGACAAAGTGATGATGGAACACTGGGATGGAGTGTGGTGTGATCGCGATCCT 780
Db 265 PheLysThrSerGlyCysTyrglyThrLeuGlyMetGluSerGlyValIleAlaAspPro 284
QY 781 CAAATAACAGCATCATCTGTCTGGAGTGGACTGACACACAGGCAAGCAAGACAGTGG 840
Db 285 GlnIleThrAlaSerSerValLeuGluThrphrAspHisThrGlyGlnGluAsnSerTrp 304
QY 841 AAACCCAAAAAGCCAGGCTGAAAAAACCCTGGACGCCCTGGCTGCTTTTGGCCACTGAT 900
Db 305 LysProLysLysAlaArgLeuLysLysProGlyProProTrpAlaAlaPheAlaThrAsp 324
QY 901 GAATACCAGTGGTTACAATAGATTGTAATAGGAAGAAAGAAATAACAGGCAATATAACC 960
Db 325 GluTyrglnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThr 344
QY 961 ACTGGATCCACCATGGTGGAGCACAAATTAATGTGTCTGCTACAGAAATCTGTACAGT 1020
Db 345 ThrGlyIleThrMetValGluHisasnTyrrValSerAlaTyrrArgIleLeuTyrrSer 364
QY 1021 GATGATGGCAGAAATGGACTGTGTACAGAGAGCCTGGTGTGGCAAGATAAGATATTT 1080
Db 365 AspAspGlyGlnLysTrpThrValTyrrArgGluProGlyValGluGlnAspLysIlePhe 384
QY 1081 CAAGGAACAAGATATTCACAGGATGGTGGTAATACTTTTCCACCAATATTGCA 1140
Db 385 GlnGlyAsnLysAspTyrrHisGlnAspValArgAsn***PheLeuProIleIleAla 404
QY 1141 CTTTTATTAGATCAATCCTACCAATGGCAGCAGAAATGGCATGAAATGGAGCTG 1200
Db 405 ArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeu 424
QY 1201 CTCGGATGTCAGTTTATTCCTTAAGGTGCTCCTCCAAAACCTTACTCAACCTCCACCTCCT 1260
Db 425 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProProPro 444
QY 1261 CGGACAGCATGACCTCAAAACACTACAGCCCTCCAAAATAGCCAAAGGTGCTGCC 1320
Db 445 ArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla 464
QY 1321 CCAAAATTTACCAACCACTCAACCTCGCAGTAGCAATGAATTTCTCCGACAGACAGAA 1380
Db 465 ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 484
QY 1381 CAAACACTGCCAGTCTCGATATCAGAATACTACCGTAACCTCAAAATGTAAACCAAGAT 1440
Db 485 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 504
QY 1441 GTAGCGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Db 505 ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 524
QY 1501 ATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Db 525 IleLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrrAsp 544
QY 1561 TTACTTACTGGGACCGGCGAGTGGTGGAAAGAAATG 1599
Db 545 LeuProTyrrTrpAspArgAlaGlyAsnSerArgGlyLeu 557
RESULT 3
AAE22716
ID AAE22716 standard; Protein: 385 AA.
XX
AC AAE22716;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy2 protein.
XX

KW Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive; chromosome 6q21.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Peptide 1..20 /label= Signal_peptide
FT Protein 61..385 /note= "Human mature neuropilin-Hy2 protein"
FT
XX WO200222815-A1.
PN 21-MAR-2002.
XX
XX 12-SEP-2001; 2001WO-US28488.
PF
XX 11-SEP-2000; 2000US-0659671.
PR 06-SEP-2001; 2001US-317902P.
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT;
PI
XX WPI; 2002-393966/42.
DR N-PSDB; AAD35994.
XX
XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
XX useful for treating neurodegenerative diseases e.g. Alzheimer's
XX disease, and for diagnosing and mapping genetic neuronal defects -
PS Claim 3; Page 138-130; 152pp; English.
XX
XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
XX neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
XX like polypeptides and polynucleotides are useful in modulating neuronal
XX growth regenerative capacity, treating neurodegenerative diseases,
XX diagnosing and mapping genetic neuronal defects and degenerative diseases
XX like Alzheimer's disease and for treating learning and memory disorders.
XX They are also useful for inducing angiogenesis, neovascularisation, as
XX well as organ growth and development e.g. heart and other tissues.
XX Antagonists of neuropilin-like polypeptides are useful for treating
XX cancers and other malignant diseases. Neuropilin is used to treat
XX platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
XX nocturnal haemoglobinuria and is used in nerve tissue growth or
XX regeneration, in wound healing, tissue repair and replacement and in
XX healing of bones, incisions and ulcers. Compositions comprising the
XX sequences of the invention are useful for treating diseases of peripheral
XX nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
XX syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
XX spinal cord disorders, head trauma and cerebrovascular diseases e.g.
XX stroke, ulcers, immune deficiencies and immune disorders, infections by
XX hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
XX mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
XX multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
XX autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
XX mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
XX inflammatory eye diseases. The nucleic acids of the invention are used in
XX gene therapy techniques. The present sequence is human neuropilin-Hy2
XX protein. Neuropilin-Hy2 gene is located on chromosome 6q21.
XX

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <OGU>
A:Cross-references: DBJ:J084068; NID:g1620006; PIDN:BAAL2210.1; PID:g1620007
A:Experimental source: CST cell
C:Comment: This protein is required for the O-acetylation of disialoganglioside sialic acid
C:Genetics:
A:Gene: ags
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homology
F:1-21/Domain: signal sequence #status predicted <SIG>
F:28-60/Domain: EGF homology <EG1>
F:68-107/Domain: EGF homology <EG2>
F:110-267/Domain: discoidin I amino-terminal homology <DNI>
F:271-427/Domain: discoidin I amino-terminal homology <DN2>

Query Match 8.7%; Score 339; DB 2; Length 427;
Best Local Similarity 26.9%; Pred. No. 1.4e-16;
Matches 128; Conservative 56; Mismatches 160; Indels 132; Gaps 19;

Qy 5 LLLLVLLLLLEDAGAGQDGC-----GHTVLG-----PESGTLTSINYPQY 47
Db 7 LAALGVLLLCAGFAASGDFCDSSLCNGGTCLMGQNDIYCLCPGEGFTGLVCNETKG 66
Qy 48 PNS-TVCEWEIRVKMGVRVIRKFGDFIEDSDSCHFNRLIYNGIGVSRTEIG-----99
Db 67 PCSNPFCFHDAKCLVTEDT--QRGDIFFEYICQCPVGSIGHCELGCS--TKLGLGGGA 123
Qy 100 -----KYCG-LGLQNMHSIESKNEITLLFMSGI-----127
Db 124 DSQISASSVYMGFLQ-----RWGPFLARLYRTGIYNAWTASSYDSKPNQVDFLRKM 177
Qy 128 HVSG-----RGFLASYSVIDKQDLTCLDTASNFLEPFSKYCPAGCLLP 172
Db 178 RVSGVMTGASRAGRAEYLFKFAVSLDGR-----FEFIQDE-----216
Qy 173 PAETSGTIPH---GYRDSPLCMAGVHAGVYNTLGGQISVVISKGIPIYESSLANNVTS 229
Db 217 ---SGTGDKFEMGNQDNNLS-----KINMFNPLEAY-----IRLPVSVCHRGCT- 259
Qy 230 VVGHLSLTFYKSGYGTIGMESGVIAADPQITASSVLEWTDHTGTQENSMKPKKARLKK 289
Db 260 ---LRPELLGCELHGCEPLGLKNTIPDSQITASS--SYKTNLRAFGWYPLHGLRDN 313
Qy 290 PGP--PNAAFATDEYQWLOIDLNEKKITGITGTSTWVEHNYYSAYRILYSDGQKWT 347
Db 314 QGKINAMTAQNSAKENWLQVLDLGTQKVTGITQARDFGHIOYVASYKVAHSDDGQWT 373
Qy 348 VYRPGVEODKIFQGNKDYHODVRNPLPPIIARFIRVNPQWQOKIAMKMLLGC 403
Db 374 VYEQGT--SKVFOGLDNNSHKKNIFEKPPMARIYVRLPSLWNRITLRLLELGC 427

RESULT 10
A36479
milk fat globule membrane protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A36479
R:Stubbs, J.D.; Lekutis, C.; Singer, K.L.; Bui, A.; Yuzuki, D.; Srinivasan, U.; Parry, C.
Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
A:Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the existence of a novel protein family
A:Reference number: A36479; MUID:91046008; PMID:2122462
A:Accession: A36479
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-463 <STU>
A:Cross-references: GB:W38337; NID:g199142; PIDN:AAA39534.1; PID:g199143
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homology
F:28-60/Domain: EGF homology <EG1>
F:68-107/Domain: EGF homology <EG2>
F:147-303/Domain: discoidin I amino-terminal homology <DNI>
F:307-463/Domain: discoidin I amino-terminal homology <DN2>

Query Match 8.3%; Score 321; DB 1; Length 463;
Best Local Similarity 39.1%; Pred. No. 3.1e-15;
Matches 63; Conservative 30; Mismatches 62; Indels 6; Gaps 3;

Qy 245 GCYGLFGMESGVIAADPQITASSVLEWTDHTGTQENSMKPKKARLKKPG--PNAAFATDEY 302
Db 307 GCLEPLGLKNNPTDSQMSASS--SYKTNLRAFGWYPLHGLRDNQGINAWTAQNSAK 364
Qy 303 QWLQIDLNEKKITGITGTSTWVEHNYYSAYRILYSDGQKWTYVYRPGVEODKIFQG 362
Db 365 EWLQVLDLGTQRTQVTGIIQTGARDFGHIOYVASYKVAHSDDGQWTVYEEQG--SSKVFQG 422
Qy 363 NKDYHQDVNRNPLPPIIARFIRVNPQWQOKIAMKMLLGC 403
Db 423 NLDNNSHKKNIFEKPPMARIYVRLPSLWNRITLRLLELGC 463

RESULT 11
T42763
coagulation factor VIII precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42763
R:Lollar, P.
submitted to the EMBL Data Library, August 1996
A:Reference number: T22269
A:Accession: T42763
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2133 <LOL>
A:Cross-references: EMBL:U49517; NID:g1511633; PID:g1511634; PIDN:AAB06705.1
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxi-
C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A;
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-2133/Product: coagulation factor VIII #status predicted <MAT>
F:23-345/Domain: ferroxidase repeat homology <FOX1>
F:402-730/Domain: ferroxidase repeat homology <FOX2>
F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query Match 8.0%; Score 311; DB 2; Length 2133;
Best Local Similarity 24.6%; Pred. No. 1.6e-13;
Matches 126; Conservative 65; Mismatches 190; Indels 131; Gaps 19;

Qy 34 ESGTLTSINYPQYTNST-VCWEIRVKMG-----RVRIKFGDFDIEDSDSCHFNYL 85
Db 1598 EQGAEPRHNFVQ--PNETRTYFWKHMAPTEDFDCWAYFSVDLEND-----1647
Qy 86 RIYNGIGVSRTEIGKYCGGLQNMHSIESKNEITLLPMSGIHVSGRGFLASYSVIDKQD 145
Db 1648 -VHSGL-----IGPL--LICRANTLNAHGRQVTV-----QEFALFFTFD---1685
Qy 146 LITCLDTASNFLPEFESKYCPAGCLLPFAEISGRIPHGYRDSPLCMAGVHA--GVVSNT 203
Db 1686 -----ETKSWYFTENVERNCRAPCHLQMED--PTLKENYR-----FHAINGYVMDT 1729
Qy 204 LGGQISVVISKGIPIYESSLANN-----226
Db 1730 LPGLV-MAQNORIRWYLLSMGSENHISHFSGHVFSYRKEEYKMYNLYPGVFETVE 1788
Qy 227 -VTSVVG-----HLSTSL---FTFKYSGCYGTIGMESGVIAADPQITASSVLEWTD 272
Db 1789 MLPSKVGWIRIECLTIGEHLQAGMSTTFLVYSKEQAPLMASGRIRDQITAS-----1841
Qy 273 HTQENSMKPKKARLKKPGPWAAPATDEYQWLOIDLNEKKITGITGTSTWVEHNYV 332
Db 1842 --GQTGWAPKARLHVSIGSNASTKDPHSWIKVVDLAPMIHIGMTQGARQKFSLSYI 1899
Qy 333 SAYRILYSDGQKWTYVYRPGVEODKIFQGNKDYHODVRNPLPPIIARFIRVNPQWQ 392
Db 1900 SQFIYMSLDGRNQSYKNGSTGILWFFGVNDASGIKHNFNFPVIARYILHPTHYSI 1959
Qy 393 KIAMKMLLGCQF----IPKGRPKLTPPPPRNSNDLKN--TTAPPKIAKGRAPKFTQP 446

Db 1960 RSTRLMELMGCDLNSCSPLGMQNKAISDQITASSHLSNFATWSPSQARLHLOGRNA 2019
QY 447 LOPRSSNEPPAQTETQASPDIRNTVTIPNVT 478
Db 2020 WRPRVS-----SAEWLQVDLQKTVKVTGIT 2045

RESULT 12
A44258
factor VIII-associated gene B hypothetical protein - human
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
R:Accession: A44258
R:Levinson, B.; Kenrick, S.; Gamel, P.; Fisher, K.; Gitschier, J.
Genomics 14, 585-589, 1992
A:Title: Evidence for a third transcript from the human factor VIII gene.
A:Reference number: A44258; MUID:93052386; PMID:1427887
A:Accession: A44258
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-216 <LEV>
A:Cross-references: GB:M90707; NID:g182316; PIDN:AAA58466.1; PID:g182317
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F.1-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>
F.57-210/Domain: discoidin I amino-terminal homology <DN2>

Query Match 7.9%; Score 308.5; DB 2; Length 216;
Best Local Similarity 37.1%; Pred. No. 7.8e-15;
Matches 66; Conservative 31; Mismatches 72; Indels 9; Gaps 3;

QY 229 SVVGHLSLFTFKSGCYGLGMSGVIAQPDQTASSVLEWTDHTGQNSWKKPKARLK 288
Db 41 RSTRLMELMGCDLNSCSPLGMQNKAISDQITASSYF-----TNMFATWSPSKARLH 95
QY 289 KPG--PPWAAFAATDEYQWLQIDLNKKKIGITGTTGSTMVEHNYVYSAYRILYSDDGQKW 346
Db 96 LOGRSNARFPQVNPKNKEWLQVDFQTKMTGVTGCVKSLTSMYVKEFLISSQDGHQW 155
QY 347 TVYRPGVEQDKIFQGNKDYHQDVVRNPLPPIIARFIRVNPQWQOKTAMKMLGCG 404
Db 156 TLFFQNG--KVKVFGQNDSTPPVNSLDPPLLTLYRIHQWHEQIALRMEVLGCE 211

RESULT 13
A47004
coagulation factor VIII precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 18-Jun-1999
R:Elder, B.; Lakich, D.; Gitschier, J.
Genomics 16, 374-379, 1993
A:Title: Sequence of the murine factor VIII cDNA.
A:Reference number: A47004; MUID:93300511; PMID:8314577
A:Accession: A47004
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2319 <ELD>
A:Cross-references: GB:L05573; NID:g192456; PIDN:AAA37385.1; PID:g192457
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F.1-19/Domain: signal sequence #status predicted <SIG>
F.23-349/Domain: ferroxidase repeat homology <FO1>
F.402-730/Domain: ferroxidase repeat homology <FO2>
F.1686-2006/Domain: ferroxidase repeat homology <FO3>
F.2007-2156/Domain: discoidin I amino-terminal homology <DN1>
F.2160-2313/Domain: discoidin I amino-terminal homology <DN2>

Query Match 7.9%; Score 308.5; DB 2; Length 2319;
Best Local Similarity 38.0%; Pred. No. 2.8e-13;
Matches 68; Conservative 30; Mismatches 72; Indels 9; Gaps 3;

QY 228 TSVVGHLSLFTFKSGCYGLGMSGVIAQPDQTASSVLEWTDHTGQNSWKKPKARL 287

Db 2143 SSIRSTRLMELMGCDLNSCSPLGMESKVISDQITASSYF-----TNMFATWSPSQARL 2197
QY 208 KKPG--PPWAAFAATDEYQWLQIDLNKKKIGITGTTGSTMVEHNYVYSAYRILYSDDGQK 345
Db 2198 HLQGRTNARFPQVNPKNKEWLQVDFQTKMTGVTGIIQGVKSLTSMYVKEFLISSQDGHG 2257
QY 346 TVYRPGVEQDKIFQGNKDYHQDVVRNPLPPIIARFIRVNPQWQOKTAMKMLGCG 404
Db 2258 WTQILYNG--KVKVFGQNDSTPPMNSLDPPLLTLYRIHQWHEQIALRMEVLGCE 2314

RESULT 14
EZHU
coagulation factor VIII precursor [validated] - human
N:Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 08-Dec-2000
C:Accession: I54318; A00525; I58059; A23584; A26174; A43986; S63527; S66445;
R:Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A:Title: Sequence of the exon-containing regions of the human factor VIII gene.
A:Reference number: I54318; MUID:93265012; PMID:1303178
A:Accession: I54318
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1921, 'S', 1923-2351 <RES>
A:Cross-references: GB:M88648; NID:g182381; PIDN:AAA52420.1; PID:g182383
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Sec
Nature 312, 330-337, 1984
A:Title: Expression of active human factor VIII from recombinant DNA clones.
A:Reference number: A00525; MUID:85061548; PMID:6438526
A:Accession: A00525
A:Molecule type: mRNA
A:Residues: 1-2351 <WOO>
A:Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179
R:Toolle, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.
S, D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A:Reference number: I58059; MUID:85061550; PMID:6438528
A:Accession: I58059
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-74, 'V', 76-1259, 'E', 1361-2351 <RE2>
A:Cross-references: GB:K01740; NID:g182802; PIDN:AAA52484.1; PID:g182803
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.;
B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.
DNA 4, 333-349, 1985
A:Title: Characterization of the polypeptide composition of human factor VIII:C and
A:Reference number: A23584; MUID:86081164; PMID:3935400
A:Accession: A23584
A:Molecule type: mRNA
A:Residues: 1-2351 <TRU>
A:Cross-references: GB:M14113; NID:g182817; PIDN:AAA52485.1; PID:g182818
R:Eaton, D.; Rodriguez, H.; Vekhar, G.A.
Biochemistry 25, 505-512, 1986
A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleava
ity.
A:Reference number: A26174; MUID:86159740; PMID:3082357
A:Accession: A26174
A:Molecule type: protein
A:Residues: 20-36;392-399, 'X', 401-402;1668-1678;1709-1722, 'D', 1723-1725;1741-1755 <E
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.
Biochemistry 31, 3315-3325, 1992
A:Title: Identification and functional importance of tyrosine sulfate residues with
A:Reference number: A42348; MUID:92207952; PMID:1554716
A:Accession: A42348
A:Molecule type: protein
A:Residues: 20-36;356-371;392-408;582-594;1668-1669, 'X', 1671;1672-1692;1693-1708;17
A:Experimental source: recombinant material from Chinese hamster ovary cells
A:Note: sequence extracted from NCBI backbone and corrected to correspond with the p
R:Fay, P.J.; Smudzin, T.M.
J. Biol. Chem. 264, 14005-14010, 1989

A:Title: Intersubunit fluorescence energy transfer in human factor VIII.
A:Reference number: A43986; MUID:89340500; PMID:2503509
A:Accession: A43986
A:Molecule type: protein
A:Residues: 'X', 517-523, 1853-1860, 'X', 1862-1864, 'X', 1866 <FAV>
R:Leyte, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens, K.; J. Biol. Chem. 266, 740-746, 1991
A:Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for A:Reference number: A56109; MUID:91093266; PMID:1898735
A:Contents: annotation; sulfation
R:Gitschier, J.; Wood, W.I.; Goralaka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar, Nature 312, 326-330, 1984
A:Title: Characterization of the human factor VIII gene.
A:Reference number: A56196; MUID:85061547; PMID:6438525
A:Contents: annotation; introns
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M. Protein Sci. 4, 740-746, 1995
A:Title: Locations of disulfide bonds and free cysteines in the heavy and light chains of A:Reference number: A56216; MUID:95338127; PMID:7613471
A:Contents: annotation; disulfide bonds
A:Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulphydryls
R:Kjalke, M.; Heding, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M. Eur. J. Biochem. 234, 773-779, 1995
A:Title: Amino acid residues 721-729 are required for full factor VIII activity.
A:Reference number: S63927; MUID:96163459; PMID:8575434
A:Accession: S63927
A:Molecule type: protein
A:Residues: 733-752; 753-759 <KJA>
R:Lind, P.; Larsson, K.; Spira, J.; Sydow-Baackman, M.; Almstedt, A.; Gray, E.; Sandberg Eur. J. Biochem. 232, 19-27, 1995
A:Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction A:Reference number: S66445; MUID:96048024; PMID:7556150
A:Accession: S66445
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1668-1685 <LIN>
C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure to C:Genetics:
A:Gene: GDB:F8C
A:Cross-references: GDB:119124; OMIM:306700
A:Map position: Xq28-Xq28
A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63 C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro A:Pathway: blood coagulation
A:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-2351/Product: coagulation factor VIII #status experimental <MAT>
F:20-740/Product: coagulation factor VIIa heavy chain #status experimental <ACH>
F:20-356/Domain: A1 <DAL>
F:23-348/Domain: ferroxidase repeat homology <FO1>
F:392-759/Domain: A2 <DA2>
F:402-730/Domain: ferroxidase repeat homology <FO2>
F:760-1667/Domain: B <DBO>
F:1668-2351/Product: coagulation factor VIIa light chain #status experimental <ACL>
F:1709-2038/Domain: A3 <DA3>
F:1716-2038/Domain: ferroxidase repeat homology <FO3>
F:2039-2191/Domain: C1 <DCL>
F:2039-2188/Domain: discoidin I amino-terminal homology <DN1>
F:2192-2351/Domain: C2 <DC2>
F:2192-2345/Domain: discoidin I amino-terminal homology <DN2>
F:60-258, 601, 776, 803, 847, 919, 962, 1020, 1024, 1074, 1085, 1204, 1274, 1278, 1301, 1319, 1403, 1 F:172-198, 267, 348, 547-573, 649-730, 1851-1877, 1918-1922, 2040-2188/Disulfide bonds: #status F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted F:365, 737, 738, 742, 1683, 1699/Binding site: sulfate (Tyr) (covalent) #status experimental F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental F:414, 426/Binding site: sulfate (Tyr) (covalent) #status predicted F:759-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental F:1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental F:2193-2345/Disulfide bonds: #status predicted

Query Match 7.9%; Score 308.5; DB 1; Length 2351;
Best Local Similarity 37.1%; Pred. No. 2.8e-13;
Matches 66; Conservative 31; Mismatches 72; Indels 9; Gaps 3;
Qy 229 SVVGHLSLTFTKSGCYGLGMSGVIAADPOITASSVLEWTDHTGOENSKPKARLK 288
Db 2176 SIRSTLRMELMGCDLNSCMPLGMSKAIQAIIASSYF-----TNFAIWSFSKARLH 2230
Qy 289 KPG--PPWAAAFATDEYQLQDLNKEKITGIITGTSTWVEHNYVYSAYRILYSDGOKW 346
Db 2231 LQGRSNARPPQVNNKELQVDFOKTMKVTGTGKSLTSMYKFEFLSSSDGQHOW 2290
Qy 347 TVYREGVEQDKIFOGNKDYHQDVNNELPPIARIRVNPTQMOOKIAMKMLGCO 404
Db 2291 TLFFONG--KVKVFOGNDSETPVNVSLDPLLRLLRYLRHPQSWVHQIALRMEVLGCE 2346
RESULT 15
JC5256
adipocyte transcription factor, AEBP1 - human
C:Species: Homo sapiens (man)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: JC5256
R:Ohno, I.; Hashimoto, J.; Shimizu, K.; Takaoka, K.; Ochi, T.; Matsubara, K.; Okubo, Biochem. Biophys. Res. Commun. 228, 411-414, 1996
A:Title: A cDNA cloning of human AEBP1 from primary cultured osteoblasts and its exp:
A:Reference number: JC5256; MUID:97079196; PMID:8920928
A:Accession: JC5256
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-845 <OHN>
A:Cross-references: DDBJ:D86479; NID:g1468942; PIDN:BAA13094.1; PID:g1468943
Query Match 5.8%; Score 224.5; DB 2; Length 845;
Best Local Similarity 31.1%; Pred. No. 6.8e-08;
Matches 52; Conservative 28; Mismatches 68; Indels 19; Gaps 4;
Qy 250 LGMESGVIAADPOITASSVLEWTDHTGOENSKPKARLKPK-----GPPWAAAFATD 300
Db 75 IGMESHRIEDNQIRASSML-----RHGLGAQRGLNMOTGATEDDDYDGAACAEEDA 126
Qy 301 EYQWLQIDLNKEKITGIITGTSTWVEHNYVYSAYRILYSDGOKWTVYREGVEQDKIF 360
Db 127 RTQWIEVDTRTTRFTGVITQGRDSSIHDDFTTFVGFSDSQTWVMYTN-GYE-EMTF 184
Qy 361 QGNKDYHQDVNNELPPIARIRVNPTQMOOKIAMKMLGCOQFIP 407
Db 185 HGNVDKDTFVLSLPEPVPVVARFIRIYPLTWNGSLCMRLEVLCGSVAP 231

Search completed: January 21, 2003, 09:54:35
Job time : 22.7699 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:38:10 ; Search time 27.9573 seconds
(without alignments)
3474.573 Million cell updates/sec

Title: US-10-060-830-3

Perfect score: 3888

Sequence: 1 MPFLFLLLVLLVLLLEDAGA.....TQEVSGAGRGDCDFEKEIL 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3572	91.9	669	21	Human Factor 8 Hom
2	2806	72.2	583	22	Human colon cancer
3	1669.5	42.9	385	23	Human neuropilin-H
4	1669.5	42.9	385	23	Human neuropilin-H
5	1669.5	42.9	385	23	Novel human protei
6	1578.5	40.6	365	23	Human neuropilin-H
7	1042	26.8	715	22	Human TANGO 229 po
8	1039	26.7	197	22	Novel human diagno
9	924.5	23.8	539	22	Novel human protei
10	924.5	23.8	586	22	Novel human protei

11	860.5	22.1	487	22	AAU00628	Novel human protei
12	574.5	14.8	503	21	AA19126	Polypeptide isolat
13	540.5	13.9	398	23	AAE22715	Human neuropilin-H
14	540.5	13.9	398	23	AAU79459	Human neuropilin-H
15	412	10.6	75	22	AB42217	Peptide #9723 enco
16	412	10.6	75	22	AA63102	Human brain expres
17	412	10.6	75	22	AA75913	Human bone marrow
18	412	10.6	75	22	AA36024	Human peptide enco
19	412	10.6	75	23	AB45349	Human peptide enco
20	404.5	10.4	957	22	AA25770	Human protein sequ
21	402.5	10.4	925	20	AA96308	Neuropilin-2. Rat
22	400.5	10.3	538	22	AAU02950	Angiotensin conver
23	400.5	10.3	600	22	AAU02948	Angiotensin conver
24	400.5	10.3	644	20	AAU06319	Human soluble neur
25	400.5	10.3	840	22	AAU02949	Angiotensin conver
26	400.5	10.3	856	22	AA62476	Human NP-1 recepto
27	400.5	10.3	923	20	AAU06317	Human neuropilin-1
28	400.5	10.3	923	20	AAU23247	Human VEGF165R/NP-
29	400.5	10.3	923	20	AA96246	Human semaphorin r
30	400.5	10.3	923	23	AAE22717	Human neuropilin-1
31	400.5	10.3	924	22	AA62476	Human neuropilin-1
32	399.5	10.3	901	20	AA96256	Mouse semaphorin r
33	399.5	10.3	906	20	AA96257	Mouse semaphorin r
34	399.5	10.3	909	20	AA96249	Rat semaphorin rec
35	399.5	10.3	909	20	AA96250	Mouse semaphorin r
36	399.5	10.3	909	21	AA624214	Mouse soluble neur
37	399.5	10.3	914	20	AA96251	Mouse semaphorin r
38	399.5	10.3	926	20	AA96252	Mouse semaphorin r
39	399.5	10.3	931	20	AA96253	Mouse semaphorin r
40	398.5	10.2	923	20	AAU14562	Human VEGF(165)R/N
41	397.5	10.2	487	22	AA64570	Human secreted pro
42	395.5	10.2	889	21	AA624216	Soluble neuropilli
43	395.5	10.2	909	20	AA96254	Mouse semaphorin r
44	395.5	10.2	926	20	AA96255	Mouse semaphorin r
45	395.5	10.2	926	23	AAE22718	Human neuropilin-2

ALIGNMENTS

RESULT 1

AAU70539
ID AAU70539 standard; Protein; 669 AA.

XX AAU70539;

AC 04-JUL-2000 (first entry)

DT Human Factor 8 Homologue.

DE Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;

KW cerebroprotective; therapeutic; coagulation related disorder;

KW haemophilia; stroke; screening.

XX Homo sapiens.

OS WO200012532-A1.

PN 09-MAR-2000.

PD 20-AUG-1999; 99WO-US19047.

XX 31-AUG-1998; 98US-0098521.

XX (ELIL) LILLY & CO ELI.

XX Rostek PRJ, Su W, Li XM;

XX WPI; 2000-256580/22.

DR N-PSDB; AAZ51872.

XX Factor 8 homolog polypeptides and nucleic acids encoding them for

PT treating coagulation related disorders such as hemophilia and stroke

xx PS Claim 3; Page 64-66; 68pp; English.

xx CC The present sequence is a human Factor 8 homologue (F8H),

xx CC a coagulation cofactor which is selectively expressed in

xx CC haematopoietic, heart and reproductive tissues. It has haemostatic and

xx CC cerebroprotective activities. The F8H contains a Factor 5/8 signature

xx CC and is useful as a therapeutic for treating coagulation related diseases

xx CC such as haemophilia and stroke. The nucleic acid is useful as

xx CC hybridisation probe and amplification primer for detecting deficiencies

xx CC in the level of F8H mRNA, for screening F8H gene mutations and for

xx CC monitoring regulation of gene expression. Fragments of the nucleic acid

xx CC are also useful as diagnostic probes and primers, and can be used in

xx CC screening methods such as those using DNA chips. The present sequence is

xx CC also useful as a target to screen therapeutically useful modulators

xx CC of the F8H.

xx	Sequence	669 AA;
xx	Query Match	91.9%; Score 3572; DB 21; Length 669;
xx	Best Local Similarity	100.0%; Pred. No. 5.2e-299;
xx	Matches	669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	61	MGERVRIKFGDFIEDSDSCHFNLYRNYGIGVSRTEIGKYCGGLQNMHSIESKGNIT 120
Db	1	MGERVRIKFGDFIEDSDSCHFNLYRNYGIGVSRTEIGKYCGGLQNMHSIESKGNIT 60
Qy	121	LLFMGSHVSGRGLASVVDKQDLTCLDTASNFLPEPFSKYCPAGCLLPFAEISGTI 180
Db	61	LLFMGSHVSGRGLASVVDKQDLTCLDTASNFLPEPFSKYCPAGCLLPFAEISGTI 120
Qy	181	PHGRDSSPLCMAGVHAGVSVNTLGGQISVVISKGIPIYESSLANNVTSVGHLSLTFT 240
Db	121	PHGRDSSPLCMAGVHAGVSVNTLGGQISVVISKGIPIYESSLANNVTSVGHLSLTFT 180
Qy	241	FKTSCYCTLGMSGVADPOITASSVLEWTDHTGQNSWPKPKARLKKPGPPWAAATD 300
Db	181	FKTSCYCTLGMSGVADPOITASSVLEWTDHTGQNSWPKPKARLKKPGPPWAAATD 240
Qy	301	EYQWLQIDLNEKKITGIITGSTMVHNYVYSAYRILYSDDGOKWTVYRPGVEQDKIF 360
Db	241	EYQWLQIDLNEKKITGIITGSTMVHNYVYSAYRILYSDDGOKWTVYRPGVEQDKIF 300
Qy	361	QGNKDYHODVRNNFLPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPPKLTOPPPP 420
Db	301	QGNKDYHODVRNNFLPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPPKLTOPPPP 360
Qy	421	RNSDLKNTTAPPKIAGRAPKFTQPLQPRSSNEFFPAQTEQTASPDIRNTVTPNVTKD 480
Db	361	RNSDLKNTTAPPKIAGRAPKFTQPLQPRSSNEFFPAQTEQTASPDIRNTVTPNVTKD 420
Qy	481	VALAAVLVPVLVMTTLILILVCAWHNRNKKKTEGTIDLPYWDRAWGMKMQFLPAK 540
Db	421	VALAAVLVPVLVMTTLILILVCAWHNRNKKKTEGTIDLPYWDRAWGMKMQFLPAK 480
Qy	541	AVDHEETPVYSSSEVNHLSPREVTTLVLOADSAEYAPLVGGIVGTLHQRTFPKPEGKE 600
Db	481	AVDHEETPVYSSSEVNHLSPREVTTLVLOADSAEYAPLVGGIVGTLHQRTFPKPEGKE 540
Qy	601	AGYADLDYNSPGQEVHAYAEPLITGPEYATPIIDMSGHPTTSVQSPSTSTFKATGN 660
Db	541	AGYADLDYNSPGQEVHAYAEPLITGPEYATPIIDMSGHPTTSVQSPSTSTFKATGN 600
Qy	661	QPPVLVGYNTLLSRTDSCSSAQAYDTPKAGKPGLPAPDELVYVQVQSTQEVSSAGRDG 720
Db	601	QPPVLVGYNTLLSRTDSCSSAQAYDTPKAGKPGLPAPDELVYVQVQSTQEVSSAGRDG 660
Qy	721	ECDFVKEIL 729
Db	661	ECDFVKEIL 669

RESULT 2

AAAG75450

xx ID AAG75450 standard; Protein; 583 AA.

xx AC AAG75450;

xx DT 03-SEP-2001 (first entry)

xx DE Human colon cancer antigen protein SEQ ID NO:6214.

xx KW Human; colon cancer; colon cancer antigen; diagnosis; detection; colorectal carcinoma.

xx OS Homo sapiens.

xx PN WO200122920-A2.

xx PD 05-APR-2001.

xx PF 28-SEP-2000; 2000WO-US26524.

xx PR 29-SEP-1999; 99US-0157137.

xx PR 03-NOV-1999; 99US-0163280.

xx PA (HUMA-) HUMAN GENOME SCI INC.

xx PI Ruben SM, Barash SC, Birse CE, Rosen CA;

xx DR WPI; 2001-235357/24.

xx DR N-PSDB; AAH34855.

xx PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers -

xx PS Claim 11; Page 7657-7660; 9803pp; English.

xx AAH32943 to AAH37195 and AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention.

xx CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for CC SEQ ID NO:1027 to 1052, 7921 and 7922.

xx SQ Sequence 583 AA;

Query Match 72.2%; Score 2806; DB 22; Length 583;

Best Local Similarity 98.9%; Pred. No. 5.1e-233;

Matches 527; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MPFLLLLLLLLLLELDAGAAQGGCGHTVLGPESGTLTSINYPQYPSNVCWEIRVK 60

Db 25 MPFLLLLLLLLLLELDAGAAQGGCGHTVLGPESGTLTSINYPQYPSNVCWEIRVK 84

Qy 61 MGERVRIKFGDFIEDSDSCHFNLYRNYGIGVSRTEIGKYCGGLQNMHSIESKGNIT 120

Db 85 MGERVRIKFGDFIEDSDSCHFNLYRNYGIGVSRTEIGKYCGGLQNMHSIESKGNIT 144

Qy 121 LLFMGSHVSGRGLASVVDKQDLTCLDTASNFLPEPFSKYCPAGCLLPFAEISGTI 180

Db 145 LLFMGSHVSGRGLASVVDKQDLTCLDTASNFLPEPFSKYCPAGCLLPFAEISGTI 204

```

QY 181 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVGHLSLST 240
Db 205 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVGHLSLST 264
QY 241 FKTSGCYGTGLMGESGVADPOITASSVLEWTDHTGOENSWKPKARLKKPGPPWAFATD 300
Db 265 FKTSGCYGTGLMGESGVADPOITASSVLEWTDHTGOENSWKPKARLKKPGPPWAFATD 324
QY 301 EYQWLQIDLNKKKKTGTTGTSTMVHEHNYVSAYRILYSDGQKWTYREPGVQDKIF 360
Db 375 EYQWLQIDLNKKKKTGTTGTSTMVHEHNYVSAYRILYSDGQKWTYREPGVQDKIF 384
QY 361 QGNKDYHODVRNFPPIIARFIRVNPQWQOKIAMKMLLGCQIPKGRPPKLQPPPP 420
Db 385 QGNKDYHODVRNFPPIIARFIRVNPQWQOKIAMKMLLGCQIPKGRPPKLQPPPP 444
QY 421 RNSDLKNTTAPPKTAKGRAPKFTOPLOPRSSNEFPAOTEQTASPDINTVTVNWKD 480
Db 445 RNSDLKNTTAPPKTAKGRAPKFTOPLOPRSSNEFPAOTEQTASPDINTVTVNWKD 504
QY 481 VALAAVLPVLVWLTLILVCAWHNRKKTEGTYDLPYWDRAWGKGM 533
Db 505 VALAAVLPVLVWLTLILVCAWHNRKKTEGTYDLPYWDRAWGKGM 557

RESULT 3
AAE22716
ID AAE22716 standard; Protein; 385 AA.
AC AAE22716;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy2 protein.
XX
KW Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopoena; memory; platelet; plastic anaemia; antinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcers; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neutropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive; chromosome 6q21.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Peptide 1..20
FT /label= Signal_peptide
FT Protein 61..385
FT /note= "Human mature neuropilin-Hy2 protein"

WO200222815-A1.
PN
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28488.
XX
PR 11-SEP-2000; 2000US-0659671.
PR 06-SEP-2001; 2001US-317902P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT;
XX
XX WPI; 2002-393966/42.
DR N-PSDB; AAD35994.

```

```

XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects
XX
PS Claim 3; Page 128-130; 152pp; English.
XX
CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.
CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat
CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
CC nocturnal haemoglobinuria and is used in nerve tissue growth or
CC regeneration, in wound healing, tissue repair and replacement and in
CC healing of bones, incisions and ulcers. Compositions comprising the
CC sequences of the invention are useful for treating diseases of peripheral
CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
CC stroke, ulcers, immune deficiencies and immune disorders, infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy2
CC protein. Neuropilin-Hy2 gene is located on chromosome 6q21.
XX
SQ Sequence 385 AA;
Query Match 42.9%; Score 1669.5; DB 23; Length 385;
Best Local Similarity 84.8%; Pred. No. 3.1e-135;
Matches 330; Conservative 11; Mismatches 37; Indels 11; Gaps 3;
QY 1 MFLFLLLLVLLLEDDAGAQGGCGHTVLGPESGTLTSINYPQTPNNTVCWEIRVK 60
Db 1 MFLFLLLLVLLLEDDAGAQGGCGHTVLGPESGTLTSINYPQTPNNTVCWEIRVK 60
QY 61 MGERVRIKPGDFDIEDSDSCHENYRIYNGIGVSRTEIGKYCGLGLQMNHSTESKGNFT 120
Db 61 MGERVRIKPGDFDIEDSDSCHENYRIYNGIGVSRTEIGKYCGLGLQMNHSTESKGNFT 120
QY 121 LLFMSGIHVSGRGFLASYSVIDKDLITCLDTASNFLPEFEFSKYCPAGCLLPFAEISGPI 180
Db 121 LLFMSGIHVSGRGFLASYSVIDKDLITCLDTASNFLPEFEFSKYCPAGCLLPFAEISGPI 180
QY 181 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVGHLSLST 240
Db 181 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVGHLSLST 240
QY 241 FKTSGCYGTGLMGESGVADPOITASSVLEWTDHTGOENSWKPKARLKKPGPPWAFATD 300
Db 241 FKTSGCYGTGLMGESGDRSSNSITVLEWTDHTGOENSWKPKKSAEKTWTALGAFATD 300
QY 301 EYQWLQIDLNKKKKTGTTGTSTMVHEHNYVSAYRILYSDGQKWTYREPGVQDKIF 355
Db 301 EYQWLQIDLNKKKKTGTTGTSTMVS-----TITMCLPTESCTVMNGRNLGCTESLVVE 355
QY 356 QDKIFOGKND-YHODVRNFPPIIARFI 383
Db 356 QDKIFOGKRIITRMVVRNFPPIIARLL 384

RESULT 4
AAU79460
ID AAU79460 standard; Protein; 385 AA.

```


DR N-PSDB; ABN32572.

XX An isolated polynucleotide for treating diseases associated with its

PT encoded polypeptide such as cancer and multiple sclerosis -

XX

PS Claim 20; SEQ ID NO 654; 509pp; English.

XX

CC The present invention provides the protein and coding sequences of 444

CC novel human proteins. These were isolated from expressed sequences tags

CC (ESTs). They can be used to stimulate cell growth, to regulate

CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth

CC e.g. in burn treatment, to regulate the immune system e.g. to treat

CC multiple sclerosis, to regulate activin or inhibin e.g. to treat

CC infertility, to regulate haemostasis or thrombolysis e.g. to treat

CC stroke and cancer, to screen for drugs, to treat inflammatory conditions

CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.

CC Parkinson's disease. The present sequence is a protein of the invention.

XX

XX Sequence 385 AA;

XX

Query Match 42.9%; Score 1669.5; DB 23; Length 385;

Best Local Similarity 84.8%; Pred. No. 3.1e-135;

Matches 330; Conservative 11; Mismatches 37; Indels 11; Gaps 3;

Qy 1 MPFLFLLLVLLLEDDAGAQGGCGTTLGPGSGTLTSINYPQTPNSVCEWEIRVK 60

Db 1 MPFLFLLLVLLLEDDAGAQGGCGTTLGPGSGTLTSINYPQTPNSVCEWEIRVK 60

Qy 61 MGERVRKFGDFDESDSCHFNRLRYNGIGVSRTEIGKYCGGLQNMHSIESKGNBIT 120

Db 61 MGERVRKFGDFDESDSCHFNRLRYNGIGVSRTEIGKYCGGLQNMHSIESKGNBIT 120

Qy 121 LFMGSHVSGRGLASVYVDKDLITCLDTASNLEPEPEPKSKYCPAGCLLPFAEISGTI 180

Db 121 LFMGSHVSGRGLASVYVDKDLITCLDTASNLEPEPEPKSKYCPAGCLLPFAEISGTI 180

Qy 181 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKGIPIYESSLANNTSVVGHLSLST 240

Db 181 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKGIPIYESSLANNTSVVGHLSLST 240

Qy 241 FKTSCTGTLGMSGVADPOITASSVLEWTDHFGQENSWPKARLKKPGPPNAAFTD 300

Db 241 FKTSCTGTLGMSGVADPOITASSVLEWTDHFGQENSWPKARLKKPGPPNAAFTD 300

Qy 301 EYQWLDLNLKKEKITGLITGSPWVHNHYVVSAYRIYSDS----CQKTVYREPGVE 355

Db 301 EYQWLDLNLKKEKITGLITGSPWVHNHYVVSAYRIYSDS----CQKTVYREPGVE 355

Qy 356 QDKIFQGNKD-YHODVRNFPPIIARFI 383

Db 356 QDKIFQGNKRITVRNFPPIIARLL 384

RESULT 6

AAE22721

XX AAE22721 standard; Protein; 365 AA.

XX

AC AAE22721;

XX

XX

DT 09-AUG-2002 (first entry)

XX

DE Human neuropilin-Hy2 mature protein.

XX

XX Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;

XX neurodegenerative disease; Alzheimer's disease; Learning; angiogenesis;

KW thrombocytopenia; memory; platelet; plastic anaemia; antiinflammatory;

KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;

KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;

KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;

KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;

KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;

KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;

KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;

KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;

KW nootropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;

KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;

XX immunosuppressive.

OS Homo sapiens.

XX

PN WO200222815-A1.

XX

PD 21-MAR-2002.

XX

XX 12-SEP-2001; 2001WO-US28488.

PF

XX 11-SEP-2000; 2000US-0659671.

PR

PR 06-SEP-2001; 2001US-317902P.

XX

XX (HYSE-) HYSEQ INC.

PA

XX

PI Tang YT;

XX

XX WPI; 2002-393966/42.

DR

XX

PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides

PT useful for treating neurodegenerative diseases e.g. Alzheimer's

PT disease, and for diagnosing and mapping genetic neuronal defects -

PS

PS Disclosure; Page 131-132; 152pp; English.

XX

CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and

CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-

CC like polypeptides and polynucleotides are useful in modulating neuronal

CC growth regenerative capacity, treating neurodegenerative diseases,

CC diagnosing and mapping genetic neuronal defects and degenerative diseases

CC like Alzheimer's disease and for treating learning and memory disorders.

CC They are also useful for inducing angiogenesis, neovascularisation, as

CC well as organ growth and development e.g. heart and other tissues.

CC Antagonists of neuropilin-like polypeptides are useful for treating

CC cancers and other malignant diseases. Neuropilin is used to treat

CC platelet disorders e.g. thrombocytopenia, plastic anaemia and paroxysmal

CC nocturnal haemoglobinuria and is used in nerve tissue growth or

CC regeneration, in wound healing, tissue repair and replacement and in

CC healing of bones, incisions and ulcers. Compositions comprising the

CC sequences of the invention are useful for treating diseases of peripheral

CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager

CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.

CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.

CC stroke, ulcers, immune deficiencies and immune disorders, infections by

CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,

CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.

CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,

CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes

CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune

CC inflammatory eye disease. The nucleic acids of the invention are used in

CC gene therapy techniques. The present sequence is human neuropilin-Hy2

CC mature protein.

XX

XX Sequence 365 AA;

XX

Query Match 40.6%; Score 1578.5; DB 23; Length 365;

Best Local Similarity 84.0%; Pred. No. 2e-127;

Matches 310; Conservative 11; Mismatches 37; Indels 11; Gaps 3;

Qy 21 OQDGGCGHTVLGPGSGTLTSINYPQTPNSVCEWEIRVKGERVIRKFGDFDESDSC 80

Db 1 OQDGGCGHTVLGPGSGTLTSINYPQTPNSVCEWEIRVKGERVIRKFGDFDESDSC 60

Qy 81 HFNVLRYNGIGVSRTEIGKYCGGLQNMHSIESKGNBITLLFMGSHVSGRGLASVSV 140

Db 61 HFNVLRYNGIGVSRTEIGKYCGGLQNMHSIESKGNBITLLFMGSHVSGRGLASVSV 120

Qy 141 IDKQDLITCLDTASNLEPEPEPKSKYCPAGCLLPFAEISGTIPHGVRDSSPLCMAGVHAGV 200

Db 121 IDKQDLITCLDTASNLEPEPEPKSKYCPAGCLLPFAEISGTIPHGVRDSSPLCMAGVHAGV 180

QY 201 SNTLGGQISVVISKIPYESSLANNTSVVGHLSLSTLFTKTSCTGCTGLMGESGVIA DP 260
 DQ 181 SNTLGGQISVVISKIPYESSLANNTSVVGHLSLSTLFTKTSCTGCTGLMGESGGRGS 240
 QY 261 QITASSVLEWTDHTGQENSWPKKARLUKPPGPPWAAATDEYOWLQIDLNKEKKITGIIT 320
 DQ 241 SNNSTVLEWTDHTGQENSWPKKARLUKPPGPPWAAATDEYOWLQIDLNKEKKITGIIT 300
 QY 321 TGSTWVEHNYVSAVRIYSD- - - - -GQKWTYVREPGVEODKIFOGNKO-YHQDVRRNF 374
 DQ 301 TGSTWVS- - - - -TTMCLPTSCVTMMGRNGLCTESLVVEODKIFOGNKRITIRWRVNF 355
 QY 375 LPPIIARFI 383
 DQ 356 LPPIIARLL 364

RESULT 7
 AAU00670
 ID AAU00670 standard; Protein: 715 AA.
 XX AC AAU00670;
 XX DT 07-SEP-2001 (first entry)
 XX DE Human TANGO 229 polypeptide.
 KW Human: TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung; skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte; bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer; anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis; attention deficit disorder; Crohn's disease; gastroenteritis; goitre; hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism; muscular dystrophy; immuno-competence; vertebrate; blood; serum.
 XX OS Homo sapiens.
 XX FH Key
 FT Peptide 1..34
 FT Domain /note= "Signal peptide"
 FT Protein /note= "Extracellular domain"
 FT Domain /note= "Mature human TANGO 229"
 FT Domain /note= "Transmembrane domain"
 FT Domain /note= "Cytoplasmic domain"
 XX WO200129088-A1.
 XX 26-APR-2001.
 XX 23-JUN-2000; 2000WO-US17386.
 XX 19-OCT-1999; 99US-0420707.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;
 DR WPI; 2001-308477/32.
 DR N-PSDB; AAS00660.
 XX New isolated nucleic acid molecule for diagnosis, prevention, and therapy of human and other animal disorder, or as modulating agent for regulating cellular processes
 PS Claim 8; Fig 1; 263pp; English.
 CC The sequence represents human TANGO 229 polypeptide. This protein and

CC similar others exhibit the ability to affect growth, proliferation, survival, differentiation, activity, morphology, or movement/migration of, e.g., T cells and cells of the heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus tissue. They can be used as modulating agents for regulating cellular processes, thus, the proteins and their associated nucleic acids can be used to prognosticate, prevent, diagnose, or treat disorders associated with physiological processes. These disorders include abnormal blood coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention deficit disorder, Crohn's disease, gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and muscular dystrophy. Antibodies to disorders such as these can be made by providing a polypeptide of the invention to an immuno-competent vertebrate and harvesting blood or serum from the vertebrate.
 XX Sequence 715 AA;
 QY Query Match 26.8%; Score 1042; DB 22; Length 715;
 DQ Best Local Similarity 36.5%; Pred. No. 1..le-80;
 XX Matches 264; Conservative 113; Mismatches 238; Indels 108; Gaps 24;
 QY 5 LLLLLLLLLLLEDAGAGCGCGHTVLGPESGTLTSINYPOTYPSNVCEWEIRVKMER 64
 DQ 20 LALLAVSAPLRQAEEELGDCGHLVYQDSGTMSTKNYPGTPNHTVCEKTIITVPKGR 79
 QY 65 VRKEGDFDIEDSDCHFNRLIYNGIVSRTEIGKVCGLGLQNHSTESKNETLLFM 124
 DQ 80 LIURLGLDDE-SOTCASDYLFTS- - -SSDQYGPYG-SMTVPKELLNTSVTVRFE 133
 QY 125 SGHVSGRGFLASVSDIKQDLITCLDTASNFLEPEFSKYCPAGCLLPFAISGTIPRGY 184
 DQ 134 SGSHISGRGFLITVASSDHPDLITCLERASHYLATEYSKFCPCAGCDVAGDISGMVDGY 193
 QY 185 RDSPLCMAGHAGVSVNTLGGQISVVISKIPYESSLANNTSVVGHLSLSTLFTKTS 244
 DQ 194 RDTSLCKAAIHAGIIADLGGQISVLQKGISRYEGLANGVLSRDSGLSKRFLFTSN 253
 QY 245 GCGYTLGMSGVIAADPOITASSVLEWTDHTGOENSWPKKARLUKPPGPPWAAATDE- - - 301
 DQ 254 GCSRLSFE- - -PDGQIRASSSWQSVNESGDQVHNSPGQARLODQGSWAGSSNNHK 309
 QY 302 -YQWLQIDLNKEKKITGIITGSTMVEHNYVSAVRIYSDDGQKWTYVREPGVEODKIF 360
 DQ 310 PREWLEIDLGEKKKITGITRTGTSQSNFNVKSVFMNKNNSKWKTKYKGVNNEKVF 369
 QY 361 QGNKDYHQDVRRNLPPIIARFIRVNPQWQOKTAMKMWELGCGFIPKGRPKLTQPPPP 420
 DQ 370 QGNSNFRDPVQNNFIPPIVARYVRVPQTWHQRIALKVELIGCQ- - - - -ITQ- - - - 416
 QY 421 RNSND- - -LKNTPAPPKIAGRAPKFTQPLQPRSSNEFFPAQTQTASPDIRMTVTPN 476
 DQ 417 -GNDSLWRKTSQTSVSTKKEDETTTRPI- - - - -PSEET- - -STGINTT- - - 458
 QY 477 VTKDVALAALVPLVLMVLTLLILVCAMHWRNKKKTEGYDLPYW- - - - -DRAGMWKG 532
 DQ 459 - - -VAIPLVLLVVLVFGMGIPAAE- - - - -RKKKKGS- - -PYGSAEAQKDCWKQ 503
 QY 533 MKQFLPAKAVDHEETP-VRIYSSE- -VNHLSPREV- - -TVLQADSABYAPLVGGVIGT 586
 DQ 504 IKY- - - - -PFARHQSAEFTISYDNEKEMTKLDLITSDMADYQOQLMIGTGTV 551
 QY 587 LHORSTFKP- -EBEGKEAGYADLDP- - -YNSPGQEVHAYAEPLITGPEYATPIIMDSG 641
 DQ 552 TRKGSTFRPMDTDAEAG-VSTDAGGHYDCFPQAGRHEIYALPLAPPEYATPIV- - -ER 607
 QY 642 HPVTSVGPQSTVFKAQGNQ- - - - -PPL- - -VGTYNTLLSRDSCSSAAQOYDT 688
 DQ 608 HVLRAHTFSAQSGYRVPGPQPGHKSLSGSGFSPVAGVAGQDGYQRPSPHQAQDRGYDR 667
 QY 689 PKA 691
 III

Db 668 PKA 670

RESULT 8

ID ABG01298 standard; Protein: 197 AA.

AC ABG01298;

XX 13-FEB-2002 (first entry)

DT DT

XX Novel human diagnostic protein #1289.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

KW Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

PF 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PR (HYSE-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

XX N-PSDB; AAS65485.

DR New isolated polynucleotide and encoded polypeptides, useful in

XX diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 20; SEQ ID NO 31657; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 197 AA;

Query Match 26.7%; Score 1039; DB 22; Length 197;

Best Local Similarity 100.0%; Pred. No. 2.8e-81;

Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 533 MKQFLPAKAVDHEETPVRYSSSEVNHLSPRETTVLQADSAEYAPLVGGVIGTLHQRT 592

DB 1 MKQFLPAKAVDHEETPVRYSSSEVNHLSPRETTVLQADSAEYAPLVGGVIGTLHQRT 60

QY 593 FKPEGKEAGYADLDPNYNSPGQEVYHAYAEPLPITGPEYATPIIMDSGHPTTSVGQPS 652

DB 61 FKPEGKEAGYADLDPNYNSPGQEVYHAYAEPLPITGPEYATPIIMDSGHPTTSVGQPS 120

QY 653 STFRATGNQPPPLVGTNTLLSRDSCSSAAQYDTPKAGKGLPAPDELVYQVQPSQOE 712

DB 121 STFRATGNQPPPLVGTNTLLSRDSCSSAAQYDTPKAGKGLPAPDELVYQVQPSQOE 180

QY 713 VSGAGRDGECDFKEIL 729

DB 181 VSGAGRDGECDFKEIL 197

RESULT 9

AAU00630

ID AAU00630 standard; Protein: 539 AA.

XX AC AAU00630;

XX 29-AUG-2001 (first entry)

DT DT

XX Novel human protein (NHP) sequence #3.

DE Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX obesity; high blood pressure; connective tissue disorder; infertility;

KW NHP-mediated pathway.

OS Homo sapiens.

XX WO200129219-A1.

PN 26-APR-2001.

XX 08-OCT-2000; 2000WO-US28798.

PF 19-OCT-1999; 99US-0160285.

XX 18-FEB-2000; 2000US-0183583.

PR (LEXI-) LEXICON GENETICS INC.

XX Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

PI WPI; 2001-290917/30.

XX N-PSDB; AAS00615.

DR Novel nucleic acid encoding human CUB-domain containing protein, useful

CC for drug screening, diagnosis and treatment of physiological disorders

CC or diseases

XX Claim 5; Page 29-30; 33pp; English.

PS The sequence represents a novel human protein (NHP) containing a CUB

CC domain (an extracellular domain). CUB proteins have been associated with

CC regulating development, modulating cellular processes and preventing

CC infectious disease. NHP nucleotide sequences are useful for gene therapy

CC of physiological disorders or diseases. NHP oligonucleotides are useful

CC as hybridisation probes for screening libraries and assessing gene

CC patterns. NHP nucleotide sequences are useful for detecting mutant or

CC inappropriately expressed NHPs (for example, those proteins associated

CC with obesity, high blood pressure, connective tissue disorders and

CC infertility) for the diagnosis of a disease. The polynucleotides may also

CC be used in screening for drugs effective in the treatment of symptomatic

CC or phenotypic manifestations of perturbing the normal function of NHP in

CC the body. Nucleotide constructs encoding NHP products are used to

CC genetically engineer host cells to express such products in vivo. These

CC host cells allow for the identification of compounds that bind to NHP

CC receptors or trigger NHP-mediated pathways.

XX

SQ Sequence 539 AA;

Query Match 23.8%; Score 924.5; DB 22; Length 539;

Best Local Similarity 37.6%; Pred. No. 9.7e-71;

Matches 221; Conservative 98; Mismatches 185; Indels 83; Gaps 17;

```

QY 5 LLLLVLLLLLEDDAGAQOQDGGHTVLGPESGTLTISINYPOTYNSTVCWEIRVKMGER 64
DB 20 LALLAVSAPLRLQAEELGDCGHLVTVQDSGTMTSKNYPGTYPNHTVCEKTIITVPKGR 79
QY 65 VRIKFGDFIEDSDSCHENFYRIYNGIGVSRTEIGKYCGGLGLOMNHSTESKGNITLLFM 124
DB 80 LILRLGLDIE-SOTCASDYLFTS-----SSDQGYPG-SMTVPKELLNTSEVTVRFE 133
QY 125 SGHVSGRGFLASVYIDKQDITCLDTASNFLEPEFSKYCPAGCLLPFAISGTHPIGY 184
DB 134 SSSHISGRGFLTYASSDHPDLITCLERASHYLKTEYKFCPCACRDVAGDISGNMVDGY 193
QY 185 RDSPLCMAGVHAGVSVNTLGGQISVWISKGIPYESSLANNVTSVVGHLSLSTLFTFKTS 244
DB 194 RDTSLCKAAIHAGIIADELGGQISVLQKRGISRYEGILANGVLSRDGSLSDKRLFTSN 253
QY 245 GCYGLGMSGVADIADPQTASSVLEWTDHTGOENSWKPKARLKKPGPPWAAAFATDE--- 301
DB 254 GCSRLSFE---PDQIRASSWSQSVNESGDQVHWSFGQARLQDQGPSWASGDSNNHK 309
QY 302 -YQWLQIDLNKEKKTITGIITGSTMVHNYSYRILYSDGOKWTVYRPGVQDKIF 360
DB 310 PREWLEIDLGEKKKTIGRTGTSGTQSNFNFYKSFVMNFKNNKWKTYKGIIVNNEKVF 369
QY 361 QGNKDYHQDVNRNFPPIIARFIRVNPQWQOKIAMKMLLGGQFIPKGRPPKLTQPPPP 420
DB 370 QGNSNFRDPVQNNFPIIVARYVRVVPQTHQRIALVELIGCQ-----ITQ----- 416
QY 421 RNSND-----LKNTPAPPKIAGRAPKFTQPLQPRSSNEFFPAQTQTTASPDIRNTVTPN 476
DB 417 -GNDSLVWRKTSQSTSVSTKKEDETIIRPI-----PSEET--STGINITT----- 458
QY 477 VTKDVALAALVPLVPLVMTLTLILLCVCAWHNRNKKKTEGYDLPYW-----DRAGWK 532
DB 459 ----VAIPLVLLVVLVFMGIFAFAF-----RKKKKGS---PYGSAQAQKTDCKWQ 503
QY 533 MKQFLPAKAVDHEETP-VRYSSSE--VNHLSPREVTTLVQADSAEA 576
DB 504 IKY-----PFAHQSAEFTISYDNEKEMTKQLDLITSDMA 538

RESULT 10
AAU00629 standard; Protein: 586 AA.
AC AAU00629;
DT 29-AUG-2001 (first entry)
DE Novel human protein (NHP) sequence #2.
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
OS Homo sapiens.
PN WO200129219-A1.
PD 26-APR-2001.
PF 08-OCT-2000; 2000WO-US28798.
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX (LEXI-) LEXICON GENETICS INC.
XX Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-290917/30.
DR N-PSDB; AAS00614.

```

```

XX
PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
XX or diseases -
XX Claim 2; Page 27-28; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
XX
SQ Sequence 586 AA;

```

```

Query Match 23.8% Score 924.5; DB 22; Length 586;
Best Local Similarity 37.6%; Pred. No. 1.1e-70;
Matches 221; Conservative 98; Mismatches 185; Indels 83; Gaps 17;

QY 5 LLLLVLLLLLEDDAGAQOQDGGHTVLGPESGTLTISINYPOTYNSTVCWEIRVKMGER 64
DB 67 LALLAVSAPLRLQAEELGDCGHLVTVQDSGTMTSKNYPGTYPNHTVCEKTIITVPKGR 126
QY 65 VRIKFGDFIEDSDSCHENFYRIYNGIGVSRTEIGKYCGGLGLOMNHSTESKGNITLLFM 124
DB 127 LILRLGLDIE-SOTCASDYLFTS-----SSDQGYPG-SMTVPKELLNTSEVTVRFE 180
QY 125 SGHVSGRGFLASVYIDKQDITCLDTASNFLEPEFSKYCPAGCLLPFAISGTHPIGY 184
DB 181 SSSHISGRGFLTYASSDHPDLITCLERASHYLKTEYKFCPCACRDVAGDISGNMVDGY 240
QY 185 RDSPLCMAGVHAGVSVNTLGGQISVWISKGIPYESSLANNVTSVVGHLSLSTLFTFKTS 244
DB 241 RDTSLCKAAIHAGIIADELGGQISVLQKRGISRYEGILANGVLSRDGSLSDKRLFTSN 300
QY 245 GCYGLGMSGVADIADPQTASSVLEWTDHTGOENSWKPKARLKKPGPPWAAAFATDE--- 301
DB 301 GCSRLSFE---PDQIRASSWSQSVNESGDQVHWSFGQARLQDQGPSWASGDSNNHK 356
QY 302 -YQWLQIDLNKEKKTITGIITGSTMVHNYSYRILYSDGOKWTVYRPGVQDKIF 360
DB 357 PREWLEIDLGEKKKTIGRTGTSGTQSNFNFYKSFVMNFKNNKWKTYKGIIVNNEKVF 416
QY 361 QGNKDYHQDVNRNFPPIIARFIRVNPQWQOKIAMKMLLGGQFIPKGRPPKLTQPPPP 420
DB 417 QGNSNFRDPVQNNFPIIVARYVRVVPQTHQRIALVELIGCQ-----ITQ----- 463
QY 421 RNSND-----LKNTPAPPKIAGRAPKFTQPLQPRSSNEFFPAQTQTTASPDIRNTVTPN 476
DB 464 -GNDSLVWRKTSQSTSVSTKKEDETIIRPI-----PSEET--STGINITT----- 505
QY 477 VTKDVALAALVPLVPLVMTLTLILLCVCAWHNRNKKKTEGYDLPYW-----DRAGWK 532
DB 506 ----VAIPLVLLVVLVFMGIFAFAF-----RKKKKGS---PYGSAQAQKTDCKWQ 550
QY 533 MKQFLPAKAVDHEETP-VRYSSSE--VNHLSPREVTTLVQADSAEA 576
DB 551 IKY-----PFAHQSAEFTISYDNEKEMTKQLDLITSDMA 585

RESULT 11
AAU00628

```

ID AAU00628 standard; Protein; 487 AA.
XX
AC AAU00628;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #1.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
XX 08-OCT-2000; 2000WO-US28798.
XX
XX 19-OCT-1999; 99US-0160285.
XX
XX 18-FEB-2000; 2000US-0183583.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX
XX WPI; 2001-290917/30.
XX
XX N-PSDB; RAS00613.
XX
XX Novel nucleic acid encoding human CUB-domain containing protein, useful
XX for drug screening, diagnosis and treatment of physiological disorders
XX or diseases -
XX
XX Claim 4; Page 26; 33pp; English.
XX
XX The sequence represents a novel human protein (NHP) containing a CUB
XX domain (an extracellular domain). CUB proteins have been associated with
XX regulating development, modulating cellular processes and preventing
XX infectious disease. NHP nucleotide sequences are useful for gene therapy
XX of physiological disorders or diseases. NHP oligonucleotides are useful
XX as hybridisation probes for screening libraries and assessing gene
XX patterns. NHP nucleotide sequences are useful for detecting mutant or
XX inappropriately expressed NHPs (for example, those proteins associated
XX with obesity, high blood pressure, connective tissue disorders and
XX infertility) for the diagnosis of a disease. The polynucleotides may also
XX be used in screening for drugs effective in the treatment of symptomatic
XX or phenotypic manifestations of perturbing the normal function of NHP in
XX the body. Nucleotide constructs encoding NHP products are used to
XX genetically engineer host cells to express such products in vivo. These
XX host cells allow for the identification of compounds that bind to NHP
XX receptors or trigger NHP-mediated pathways.
XX
SQ Sequence 487 AA;

Query Match 22.1%; Score 860.5; DB 22; Length 487;
Best Local Similarity 37.2%; Pred. NO. 2.8e-65;
Matches 206; Conservative 95; Mismatches 170; Indels 83; Gaps 17;

QY 38 LTSINYPOTYNSTVCEWEIRVKMKERVIRKFGDFDIEDSDSCHFNLYRIYNGIGVSRTE 97
DB 1 MTSKNYPGTYPNHTVCEKTTVPKRLILRLGLDLIE-SQTCASDYLLFTS-----SSDQ 55

QY 98 IGKYGGLGLQNMHNSYESKGNETTLFLMSGIHVSGRGFLASYSVIDKQDILCTLDASNFL 157
DB 56 YGPGYCG-SMTVPKELLNTSEVTVFESGSHISGRGFLTYASSDHPDLITCLERASHYL 114

QY 158 EPEFSKYCPAGCLLPFAEISGTPHGYRDSPLCMAGVHAGVYVNTPLGQISVTSKGP 217
DB 115 KTEYSKFCPAGCRDVGADISGNMVDGYRDTSLCKAAIHAGIADBLGQISVLRKQGIS 174
QY 218 YVESLIANNVTSVGHLSTSLFTFTSGCYGLGMESGVIAQDPQITASSVLEWTDHTGQE 277

Db 175 RYEGILANGVLSRGLSDKRLFTSNGCSRLSFE----PDGQIRASSSQSVNESGDQ 230
QY 278 NSWPKPKARLKKPGPPWAAAFATDE----YQWLQIDLNKKKITGITTTGSTMVHNHYVS 333
Db 231 VHWSPGARLQDOGFPWASGSDSSNNHKKPREWLEIDLGEKKKITGITTTGTSQSNFVVK 290
QY 334 AYRILYSDGQKWTYVREPVEQDKIFOGNKDYHODVRNNEFLPPIARFIRVNPQWQOK 393
Db 291 SFVMNFKNNKWKYTKGIVNNEEKVFGNSNFRDPQVQNNFPPIVARYVRYVQTMHOR 350
QY 394 TAMKVELLGCQFIPKGRPPKLTQPPPRNSND---LKNTPAPPKIAKGRAPKETQPLQ 449
Db 351 TALKVELIGCC-----ITQ-----GNDLSVWRTSQSTSVSTKKEETIRPI-- 393
QY 450 RSSNEFFPAQTEQTASPDIRNTVTPTNVTQKVALAVLPVLMVLTLLILVCAWHWR 509
Db 394 -----PSEET--STGINITT-----VAIPLVLLVLFVAFAGMGIFAFAF----- 428
QY 510 NRKKTTEGYDLPYW----DRAGWKMGKQFLPAKAVDHEETP-VRYSSSE--VNHLSPR 562
Db 429 -RKKKKKGS---PYGSAEAQKTDCKWKQIKY-----PFARHQSAAETIISYDNEK 472
QY 563 EVTVTLQADSAEYA 576
Db 473 EMTQKLDLITSDMA 486

RESULT 12
AAB19126
ID AAB19126 standard; Protein; 503 AA.
XX
AC AAB19126;
XX
DT 19-FEB-2001 (first entry)
XX
DE Polypeptide isolated from lymph node stromal cells of fsn +/- mice.
XX
KW Lymph node stromal cell; fsn +/- mice; inflammatory disorder;
KW immune system disorder; cancer; viral infection; HIV infection;
KW blood vessel growth; tumour necrosis factor disorder; arthritis;
KW inflammatory bowel disease; fibroblast growth factor-mediated disorder;
KW cardiac failure.
XX
OS Mus sp.
XX
XX WO200058463-A1.
XX
XX 05-OCT-2000.
XX
XX 18-FEB-2000; 2000WO-NZ00015.
XX
XX 25-MAR-1999; 99US-0276268.
XX
XX 26-AUG-1999; 99US-0383586.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;
PI Murilson JG;
XX
XX WPI; 2000-664924/64.
XX
XX N-PSDB; AAA96736.
XX
XX Polypeptide expressed in mammalian fsn +/- lymph node stromal cells,
XX useful for modulating growth of blood cells, for treating inflammatory
XX and tumour necrosis factor-mediated disorders, cancer and viral
XX disorders -
XX
XX Claim 1; Page 68-69; 75pp; English.
XX
XX The present sequence represents a polypeptide sequence which is
XX isolated from lymph node stromal cells of fsn +/- mice. The
XX polynucleotides and their polypeptides are useful for treating an
XX inflammatory disorder, disorder of immune system and cancer selected

CC from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
CC viral disorder, in particular HIV infection and for modulating the
CC growth of blood vessels. The polypeptides are useful for treating a
CC tumour necrosis factor (TNF) mediated disorder, such as those selected
CC from arthritis, inflammatory bowel disease and cardiac failure and a
CC fibroblast growth factor-mediated disorder. It is also useful in assays
CC to determine biological activity, to raise antibodies, to isolate
CC corresponding ligands or receptors, to quantify levels of protein or
CC cognate corresponding ligand or receptors, as antiinflammatory agents,
CC and in compositions for the treatment of skin, connective tissue and
CC immune system diseases. The polynucleotide is useful as marker for
CC tissue, as a chromosome marker or tags in the identification of a
CC genetic disorder.
XX
SQ Sequence 503 AA;
Query Match 14.8%; Score 574.5; DB 21; Length 503;
Best Local Similarity 24.6%; Pred. No. 1.4e-40;
Matches 184; Conservative 84; Mismatches 179; Indels 301; Gaps 21;
QY 2 PFLLLLVLVLLLEDAGAQGDGGCHTVLGPESGTLTSINYPQTPNSTVCEWEIRVKM 61
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 8 PSLVALLFAVCAPLRQLQAEELDGGCHIVTSODSGTMTSKNPGYPTVCEKIITVPK 67
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 62 GERVIKFGDFDIEDSDSCHFNLYRLIYNGIGVSRTEIGKYCGLGLQMNHSTESKNEITL 121
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 68 GRLRLRLGDLNIE-SKTCASDYLFFSS-----ATDQGPYCG-SWAVPKELRLNSNEVT 121
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 122 LFMSGTHVSGRFLASYSVIDKQDLITCLDTASNFLEPFESKYCPAGLLPPAESGTP 181
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 122 LFKSGSHISGRFLITYASSDRPDLITCLGRSHVFEKYKFCPAGCRDARDISGNTK 181
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 182 HGYRSSPFCMAGVHAGVSVNTLGGQISVWISKGPYYESSLANNVTSVGHLSLSTLF 241
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 182 DGYRDTSLKAAIHAGIITDELGHINLQSGISHVEGLLANGVLSRHGSLSEKRLF 241
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 242 KTSQCYGTLGMESSGVIADPQITASSVLENTDHTGQENSWKPKKRLKPGPPWAFATDE 301
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 242 TT----- 243
QY 302 YOWQLDLNKEKKITGIITGTSTVHNVYSAYRILYSDDGQKWTVYREPGVEQDKIFQ 361
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 362 GNKDYHQDVRRNFPPIIARFIRVNPTQWQKIAMKMLLGCQIFPKRPPKLTQPPPPR 421
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 247 ----- 246
QY 422 NSNDLKNTPPKIAKRAPKFTQLPQRSNNEFFAQTEQTTASPDINTVTVPNTKDV 481
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 482 ALAAVLVPLVNLVTLILVCAWHNRKKKTEGTYDLPY---WDRAGWKGKMQFL 537
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 252 AIPSVI---FIALLTGMIFAIC-----RRKKGN---PYVSADAQKTCWQIKY-- 298
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 538 PAKAVDHEETP-VRYSSE--VNHLSPREVT---TVLQADSAXEYAPLVGGVIGTLHORS 591
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 592 IFKP-----EE-----GKAGVADLPNPSQGVYHAYAEPLTGPYPATPII----- 636
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 349 TFRPMDTTEYRVNTEASGH-----YDCPHRPGRHEVALPLTHSEYATPIVERHLRL 403
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 637 ----MDMSGHPTSVGQPSSTFKATGNQPPPLVGTNTLLSRUDSCSSAAQYDTPKAG 692
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 404 AHFTSQGYRVPG-PRTHESHSGSGFFPATGATQVESYQRPASPKPVGGGYDKP-AA 461
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 693 RGPLPAPDELVYVQSQVEYSGAGRG 720
Db | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 462 SSFLDSRD-----PASQSQMTSGGDDG 483

RESULT 13
AAE22715
ID AAE22715 standard; Protein; 398 AA.
XX
AC AAE22715;
XX
DT 09-AUG-2002 (first entry)
XX Human neuropilin-Hy1 protein.
XX
KW Human; neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW human immunodeficiency virus; HIV; autoimmune disorder; multiple sclerosis;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neurotropic; neuroprotective; vulnerrary; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive; chromosome 6q21.
XX
OS Homo sapiens.
XX WO200222815-A1.
XX 21-MAR-2002.
XX 12-SEP-2001; 2001WO-US28488.
XX 11-SEP-2000; 2000US-0659671.
XX 06-SEP-2001; 2001US-317902P.
XX (HYSE-) HYSEQ INC.
XX Tang YT;
XX WPI: 2002-393966/42.
XX N-PSDB: AAD35992.
XX
PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects
XX
PS Claim 3; Page 123-125; 152pp; English.
XX
CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.
CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat
CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
CC nocturnal haemoglobinuria and is used in nerve tissue growth or
CC regeneration, in wound healing, tissue repair and replacement and in
CC healing of bones, incisions and ulcers. Compositions comprising the
CC sequences of the invention are useful for treating diseases of peripheral
CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
CC stroke, ulcers, immune deficiencies and immune disorders, infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune

CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hyl
CC protein. Neuropilin-Hyl gene is located on chromosome 6q21.
XX
XX
SQ Sequence 398 AA;
Query Match 13.9%; Score 540.5; DB 23; Length 398;
Best Local Similarity 32.8%; Pred. No. 8.2e-38;
Matches 138; Conservative 55; Mismatches 117; Indels 111; Gaps 11;
QY 22 QGDCGHTVLGPESGTLTSINYPOTYPSNVCWEIRVKMGRIKFGDFDIEDSDSCH 81
DB 3 EGDGCHLVTYQDSGTMTSKNYPGTYPNHTVCKTITVPKGRLLRLGLDLIE-SQTCA 61
QY 82 FNYLRYINGVSRTEIKYKGLGLQMNHSIESKGNITLLFMSGIHVSGRGFLASYSVI 141
DB 62 SDYLLFTS-----SSDQYGPYCG-SMTVPKELLNTSEVTVRFESGSHISGRGFLTYASS 116
QY 142 D-----KODLTCTDTSNLFEPESKYCPAGCLLPFAEISGTIPHGYS 187
DB 117 DHPESQDRPSEKTLDDQSRFTLATGTFVKDSFS-----TDGT----- 155
QY 188 SPLCMAGVHAGVSNLTGGQISVLSKIGPIYESSLANNVTS----- 229
DB 156 SLCKAAIHAGIIADELGGQISVLQKGISRYEGILANGVLSRFEIFREQLFSSVLFS 215
QY 230 -----VVGH----- 246
DB 216 WNTVHAVIELMFPHMVYHWSKTRREGSIAAEFEGPKLVILVIOKQLVQDLVLVATVGC 275
RESULT 14
AAU79459
ID AAU79459 standard; Protein; 398 AA.
XX
XX AC AAU79459;
XX
XX DT 15-JUL-2002 (first entry)
XX
XX DE Human Neuropilin-Hyl.
XX
XX KW Human; neuropilin-Hyl; chromosome 6q21; neuronal growth;
KW nerve regeneration; neurodegenerative disease; learning disorder;
KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
KW organ growth; nervous system lesion; cancer; cell proliferation;
KW cell differentiation; stem cell growth factor activity;
KW Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;
KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
KW platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
KW reperfusion; food supplement; DNA microarray.
XX
XX OS Homo sapiens.
XX
XX PN WO200222780-A2.
XX
XX PD 21-MAR-2002.
XX
XX PF 11-SEP-2001; 2001WO-US28590.
XX
XX PR 11-SEP-2000; 2000US-0659671.
XX
XX PR 06-SEP-2001; 2001US-0659671.

XX
PA (TANG/) TANG T Y.
XX
PI Tang TY;
XX
XX WPI; 2002-351881/38.
DR N-PSDB; ABK49565.
XX
XX New neuropilin-like polypeptides for diagnosing, preventing and
PT treating neurological conditions and disorders, cancers, and for
PT inducing angiogenesis and neovascularisation -
XX
PS Claim 3; Page 118-120; 144pp; English.
XX
XX The invention relates to an isolated neuropilin-like polypeptide
CC including neuropilin-Hyl and neuropilin-Hy2, the full length cDNAs
CC encoding the proteins and the coding regions of the cDNAs. Also included
CC is a nucleic acid array comprising the cDNAs attached to a surface used
CC for detecting full-matches or mismatches to the cDNAs. The genes
CC for neuropilin-Hyl and Hy2 are located on human chromosome 6q21.
CC The nucleic acid array is useful for detecting full-matches or mismatches
CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
CC in modulating neuronal growth, regenerative capacity, treating
CC neurodegenerative diseases, learning and memory disorders, diagnosing and
CC mapping genetic neuronal defects and degenerative diseases like
CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
CC and organ growth and development (e.g. the heart). The nervous system
CC disorders include lesions of central or peripheral nervous systems,
CC including traumatic lesions, ischaemic lesions, infectious lesions,
CC degenerative lesions, lesions associated with nutritional diseases or
CC disorders, neurological lesions, and lesions caused by toxic substances.
CC The neuropilin-like proteins and cDNAs are also useful as markers for
CC cancers. The neuropilin-like proteins are useful for regulating cell
CC proliferation, cell differentiation, stem cell growth factor activity,
CC for inducing proliferation of neural cells, regeneration of nerve and
CC brain tissue, for treatment of central and peripheral nervous system
CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopaenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues and as a food supplement or molecular
CC weight marker. The cDNAs are useful in gene identification, genome
CC mapping, transgenics, as hybridisation probes, for primer design, for
CC gene chips and as a DNA antigen. The present sequence represents
CC neuropilin-Hyl.
XX
SQ Sequence 398 AA;
Query Match 13.9%; Score 540.5; DB 23; Length 398;
Best Local Similarity 32.8%; Pred. No. 8.2e-38;
Matches 138; Conservative 55; Mismatches 117; Indels 111; Gaps 11;
QY 22 QGDCGHTVLGPESGTLTSINYPOTYPSNVCWEIRVKMGRIKFGDFDIEDSDSCH 81
DB 3 EGDGCHLVTYQDSGTMTSKNYPGTYPNHTVCKTITVPKGRLLRLGLDLIE-SQTCA 61
QY 82 FNYLRYINGVSRTEIKYKGLGLQMNHSIESKGNITLLFMSGIHVSGRGFLASYSVI 141
DB 62 SDYLLFTS-----SSDQYGPYCG-SMTVPKELLNTSEVTVRFESGSHISGRGFLTYASS 116
QY 142 D-----KODLTCTDTSNLFEPESKYCPAGCLLPFAEISGTIPHGYS 187
DB 117 DHPESQDRPSEKTLDDQSRFTLATGTFVKDSFS-----TDGT----- 155
QY 188 SPLCMAGVHAGVSNLTGGQISVLSKIGPIYESSLANNVTS----- 229
DB 156 SLCKAAIHAGIIADELGGQISVLQKGISRYEGILANGVLSRFEIFREQLFSSVLFS 215
QY 230 -----VVGH----- 246
DB 216 WNTVHAVIELMFPHMVYHWSKTRREGSIAAEFEGPKLVILVIOKQLVQDLVLVATVGC 275

QY 247 YGTGLMESGVIADPQITASSVLEWTDHTQENSWKPKARLKKPGPPWAAAFATDE----Y 302
Db 276 SRSLSE---PDGQIRASSWQSVNESGDVHWSPGQARLQDQGPSWASGDSNNHKPR 331
QY 303 QWLQIDLNKKKITGITITGSTMVHEHNYVYSAYRILYSDDGOKWTVYREPGEQDKIFQG 362
Db 332 EWLEIDLGEKKKITGITITGSTQSNFNFYVKSFMNKNNSKWKTYKGIYVNEEKV-RG 390
QY 363 N 363
Db 391 N 391

RESULT 15

ABB42217
ID ABB42217 standard; Peptide; 75 AA.

XX
AC ABB42217;

XX DT 04-FEB-2002 (first entry)

XX DE Peptide #9723 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX PN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000669.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human fetal liver -

XX PS Claim 27; SEQ ID NO 34852; 639pp + sequence listing; English.

XX CC The invention relates to a single exon nucleic acid probe for

XX CC measuring human gene expression in a sample derived from human foetal

XX CC liver. The single exon nucleic acid probes may be used for predicting,

XX CC measuring and displaying gene expression in samples derived from human

XX CC fetal liver. The present sequence is a peptide encoded by a single exon

XX CC nucleic acid probe of the invention.

XX CC Note: The sequence data for this patent did not form part of the

XX CC printed specification, but was obtained in electronic format directly

XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 75 AA;

Query Match 10.68; Score 412; DB 22; Length 75;
Best Local Similarity 100.0%; Pred. No. 8.1e-28;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 DCGGHTVLGPESGTLTSINYPQTPNSTVCEWIRKMGERVRIKFGDFIEDSDSCHFN 83

Db 1 DCGGHTVLGPESGTLTSINYPQTPNSTVCEWIRKMGERVRIKFGDFIEDSDSCHFN 60

Qy 84 YLRIYNGIGVSRTEI 98

Db 61 YLRIYNGIGVSRTEI 75

Search completed: January 21, 2003, 09:52:39
Job time : 29.9573 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 01:52:19 ; Search time 72.8638 Seconds
(without alignments)
9217.500 Million cell updates/sec

Title: US-10-060-830-2

Perfect score: 2190

Sequence: 1 atgcctctgtctctctgct.....ttttaagaagaatccttga 2190

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
c 1	544.6	24.9	640	4	US-09-328-111-480
2	91.6	4.2	6585	3	US-08-746-111-4
3	86.6	4.0	6909	2	US-08-804-196-1
4	86.6	4.0	6909	2	US-08-558-340-1
5	86.6	4.0	6909	3	US-08-746-111-26
6	86.4	3.9	2303	2	US-08-480-229C-9
7	86.4	3.9	2303	2	US-08-659-235C-9
8	78.6	3.6	318	2	US-08-480-229C-19
9	78.6	3.6	318	2	US-08-659-235C-19
10	78.2	3.6	1780	2	US-08-480-229C-11
11	78.2	3.6	1780	2	US-08-659-235C-11
12	76.8	3.5	1384	1	US-07-607-538C-1
13	76.8	3.5	1384	1	US-08-162-402B-1
14	76.8	3.5	1934	2	US-08-162-402B-7
15	63.4	2.9	4629	2	US-08-484-891-7
16	63.4	2.9	4670	3	US-08-717-234-41
17	63.4	2.9	4999	4	US-08-470-618-14
18	63.4	2.9	4999	4	US-09-364-862-14
19	63.4	2.9	5035	2	US-08-882-083-1
20	63.4	2.9	5035	2	US-08-558-107-1
21	63.4	2.9	5035	3	US-09-243-539-1
22	63.4	2.9	6999	1	US-08-276-594A-1
23	63.4	2.9	7056	1	US-08-121-202-1
24	63.4	2.9	7493	1	US-08-212-133A-7
25	63.4	2.9	7493	1	US-08-474-503-5
26	63.4	2.9	7493	2	US-08-670-707A-5
27	63.4	2.9	7493	4	US-09-037-601-5

28	63.4	2.9	7493	4	US-09-315-179-5	Sequence 5, Appl
29	63.4	2.9	7493	5	PCT-US94-13200-5	Sequence 5, Appl
30	63.4	2.9	8241	6	5171844-1	Patent No. 5171844
31	63.4	2.9	9009	1	US-08-366-851A-1	Sequence 1, Appl
32	63.4	2.9	9009	1	US-07-864-004B-3	Sequence 3, Appl
33	63.4	2.9	9009	1	US-08-251-937A-3	Sequence 3, Appl
34	63.4	2.9	9009	1	US-08-212-133A-1	Sequence 1, Appl
35	63.4	2.9	9009	1	US-08-474-503-1	Sequence 1, Appl
36	63.4	2.9	9009	2	US-08-670-707A-1	Sequence 1, Appl
37	63.4	2.9	9009	4	US-09-037-601-1	Sequence 1, Appl
38	63.4	2.9	9009	4	US-09-315-179-1	Sequence 1, Appl
39	63.4	2.9	9009	4	US-09-523-656-1	Sequence 1, Appl
40	63.4	2.9	9009	5	PCT-US93-03275-3	Sequence 3, Appl
41	63.4	2.9	9009	5	PCT-US94-13200-1	Sequence 1, Appl
42	63.4	2.9	9354	1	US-08-683-839B-2	Sequence 2, Appl
43	63.4	2.9	11933	4	US-09-470-618-13	Sequence 13, Appl
44	63.4	2.9	11933	4	US-09-364-862-13	Sequence 13, Appl
45	63.2	2.9	7032	4	US-09-324-867-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-328-111-480/c
; Sequence 480, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Carrino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
; US-09-328-111-480

Query Match 24.9%; Score 544.6; DB 4; Length 640;
Best Local Similarity 96.7%; Pred. No. 2.9e-165;
Matches 616; Conservative 0; Mismatches .15; Indels 6; Gaps 6;

Qy 1121 TTTTCCACCATAATATTCACGCTTTTATAGAGTGAATCCCTACCAATGGCAGCAGAAAA 1180
Db 632 TTTTCCACCATAATATTCACGCTTTTATAGAGTGAATCCCTACCAATGGCAGCAGAAA 576
Qy 1181 TTGCATGAATGAAGTGTCTCGATGTCAGTTTATTCCTAAAGTGTCTCTCAAAAC 1240
Db 575 TTGCATGAATGAAGTGTCTCGATGTCAGTTTATTCCTAAAGTGTCTCTCAAAAC 518
Qy 1241 TTACTCAACCTCCACCTCCTCGGAACAGCAATGACCTCAAAACACTACAGCCCTCCAA 1300

Db	517	TTACTCAACTCCACCTCTTNGNNAACAGCAATGNCCTCAAAACACACTACAGCCCTCC-A	459
QY	1301	AAATAGCCAAAGGTCGTGCCCCAAAATTTAGCGAACCACTACAACCTCGCAGTAGCAATG	1360
Db	458	AAATAGCCAAAGGTCGTGCCCCAAAATTTAGCGAACCACTACAACCTCGCAGTAGCAATG	399
QY	1361	AATTTCTCTGCAGACAGACAAACAACACTGCCAGTCTCGATATCAGAAATACTACCGTAA	1420
Db	398	AATTTCTCTGCAGACAGACAAACAACACTGCCAGTCTCGATATCAGAAATACTACCGTAA	339
QY	1421	CTCCAAATGTAACCAAGATGTAGCGTGGCTGCAGTTCTTTGTCCTCTGCTGGTCATGG	1480
Db	338	CTCCAAATGTAACCAAGATGTAGCGTGGCTGCAGTTCTTTGTCCTCTGCTGGTCATGG	279
QY	1481	TCTCTACTACTCTCATTTCTCATATTAGTGTGCTTGGCACTGGAGAAACAGAAAGAAA	1540
Db	278	TCTCTACTACTCTCATTTCTCATATTAGTGTGCTTGGCACTGGAGAAACAGAAAGAAA	219
QY	1541	AAACTGAAGGCACCTATGACTTTACCTTACTGGGACGGCAGGTTGGTGGAAAGGAATGA	1600
Db	218	AAACTGAAGGCACCTATGACTTTACCTTACTGGGACGGCAGGTTGGTGGAAAGGAATGA	159
QY	1601	AGCAGTTTCTTCTGCAAAAGCAGTGGACCATAGAGAAACCCAGTTTCGCTATAGACGA	1660
Db	158	AGCAGTTTCTTCTGCAAAAGCAGTGGACCATAGAGAAACCCAGTTTCGCTATAGACGA	99
QY	1661	GCGAAGTTTAATCACCTGTAGTCCAAGAGAAGTCACACAGTCTGCAGGCTGACTCTGCAG	1720
Db	98	GCGAAGTTTAATCACCTGTAGTCCAAGAGAAGTCACACAGTCTGCAGGCTGACTCTGCAG	39
QY	1721	AGTATGCTCAGGCCTGGTAGGAGAAATTTGGTAC	1757
Db	38	AGTATGCTCAGGCCTGGTAGGAGAAATTTGGTAC	2

RESULT 2

```

US-08-746-111-4
; Sequence 4, Application US/08746111
; Patent No. 6066778
; GENERAL INFORMATION:
; APPLICANT: Ginsburg, David
; APPLICANT: Cui, Jisong
; TITLE OF INVENTION: Compositions And Methods For Screening
; TITLE OF INVENTION: Compounds For Anticoagulant Activity
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,111
; FILING DATE: 06-NOV-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolla, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: 08-02536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6585 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6..6554
; US-08-746-111-4

Query Match 4.2%; Score 91.6; DB 3; Length 6585;
Best Local Similarity 53.3%; Pred. No. 1.4e-18;
Matches 193; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 878 CTGTGGGCTGCTTTTGGCCACTGATGAATACCAAGTGGTGTACAAATAGATTTGAATAAGGAAA 937
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6214 CTTGGCAAGCCAGGCAAAACAACAAGCAGTGGTTACAAAGTCGATCTGCTCAAAATCA 6273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 938 AGAAATAACAGGCATTATTAACCACTGGATCCACCATTGGTGAGCACAATTTACTATCTGT 997
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6274 AGAAGTAAGGCCATCGTAAACGAGGCTGTAAAGTCTCTGCTCTGAGATGTACGTGA 6333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 998 CTGCTACAGAACTCTGTACAGTGTATGGCGAGAATGGACTGTGTACAGAGAGCGCTG 1057
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6334 AGAGCTACAGCATCCACTACAGTGACCCAGGCTGTGGCATGGAACCTTTACCGACAGAAAT 6393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1058 GTGTGGAGCAAGATAAGATATTTCAGGAACAAGAAGATTTATCACCGAGATGTGGGTAAATA 1117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6394 CTTCCATGGTGGCAAGATTTTGAAGGAACAACGAATACCAAGGGGCATGAAGAACT 6453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1118 ACTTTTGGCCACCAGTATTCACAGCTTTTATTAGAGTGAATCCCTACCAATGGCGAGAGA 1177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6454 TTTTCAACCGCCCATTTATTCAGATTTATCCGCATCATTCCTAAAACATGGGAACCGAGA 6513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1178 AAATTGGCATGAAATGGAGCTGCTCGGATGTCAGTTTATTCTTAAAGGTCGTCTCTCAA 1237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6514 GCATCGCCCTCGCCTAGAGCTCTTCGCGCTGTGACATTTATTAGAAATTAATTCCTCAAAA 6573
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1238 AA 1239
   ||
Db 6574 AA 6575

RESULT 3
US-08-804-196-1
; Sequence 1, Application US/08804196
; Patent No. 5874256
; GENERAL INFORMATION:
; APPLICANT: Bertina, Rogier
; APPLICANT: Reitsma, Pieter
; TITLE OF INVENTION: A method for diagnosing an increased
; TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
; TITLE OF INVENTION: thrombosis and kit for use with the same.
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5874256el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,196
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/454,353
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409

```

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 948-7400
 TELEFAX: (301) 948-9751
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6909 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: unknown
 MOLECULE TYPE: cDNA
 US-08-804-196-1

Query Match 4.0%; Score 86.6; DB 2; Length 6909;
 Best Local Similarity 51.7%; Pred. No. 6.1e-17;
 Matches 259; Conservative 0; Mismatches 224; Indels 18; Gaps 2;

```

Qy 721 TTTAAGACAAGTGGATGTTATGGAACACTGGGAGTGGTGTGATCGCGGATCCT 780
Db 6271 TGTGAGGTAATGATGTTCCACACCCCTGGGTATGGAATGGAAGATAGAAAACAAG 6330
Qy 781 CAATAACAGCATCATCTGTGCTGGAGTGACTGACCACACAGGCGCAAGACAGATTGG 840
Db 6331 CAATACACAGCTCTTC-----GTTTAAGAAATCTTGGTGGGAGATTACTGG 6378
Qy 841 AAACCCAAAAGCCAGGCTGAAAAAACCTGGAC-----CGCCTTGGGCTGCTTTTGC 894
Db 6379 GAACCCCTCCGTCGCCGCTCTGAATGCCAGGACGCTGTGAATGCTGGCAAGCCAAGCA 6438
Qy 895 ACTGATGAATACAGTGGTTACAAATAGATTGTAATAGGAAAAGAAAATACAGGCAATT 954
Db 6439 AACAAATAAGCAGTGGCTAGAAATTTGATCTACTCAAGATCAAGAAGATAACGGCAATT 6498
Qy 955 ATACCACTGGATCCACCATGCTGGAGCACAATTTACTATGTCTGCTACAGAACTCTG 1014
Db 6499 ATACACAGGGCTCCTGCTCTCTGAAATGTATGTAAGAGCTATACCATCCAC 6558
Qy 1015 TACAGTGATGATGGGAGAAATGAGTCTGTACAGAGAGCGCTGTGTGGAGCAAGATAAG 1074
Db 6559 TACAGTGAGCAGGAGTGGATGGAATGGAACCATACAGGCTGAAATCCTCCATGGTGGACAAG 6618
Qy 1075 ATATTTCAAGAAACAAGATTATACACAGGATGTCGTAATACTTTTCCACCAATT 1134
Db 6619 ATTTTGAAGGAAATACTAATACCAAGGACATGTGAAGAACTTTTCAACCCCAATC 6678
Qy 1135 ATTCACGTTTTATTAGAGTGAATCCTACCAATGGCAGCAGAAAATGGCCATGAAAATG 1194
Db 6679 ATTTCCAGGTTTATCCGTGTCATTCCTTAAACATGGAATCAAAGTATTACACTTCGCCCTG 6738
Qy 1195 GAGCTGCTCGGATGTCAGTTT 1215
Db 6739 GAACCTTTGGCTGTGATATT 6759

```

RESULT 4
 US-08-658-340-1
 Sequence 1, Application US/08658340
 Patent No. 5910576
 Patent No. 5910576 5861489

GENERAL INFORMATION:
 APPLICANT: Bertina, Rogier
 APPLICANT: Reitsma, Pieter
 TITLE OF INVENTION: A method for diagnosing an increased
 TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
 TITLE OF INVENTION: thrombosis and kit for use with the same.
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESS: Akzo No. 5910576 5861489el Patent Department
 STREET: 1300 Piccard Drive, Suite 206
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/658,340
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/454,353
 FILING DATE: 06-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Gornley, Mary E.
 REGISTRATION NUMBER: 34,409
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 948-7400
 TELEFAX: (301) 948-9751
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6909 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: unknown
 MOLECULE TYPE: cDNA
 US-08-658-340-1

Query Match 4.0%; Score 86.6; DB 2; Length 6909;
 Best Local Similarity 51.7%; Pred. No. 6.1e-17;
 Matches 259; Conservative 0; Mismatches 224; Indels 18; Gaps 2;

```

Qy 721 TTTAAGACAAGTGGATGTTATGGAACACTGGGAGTGGTGTGATCGCGGATCCT 780
Db 6271 TGTGAGGTAATGATGTTCCACACCCCTGGGTATGGAATGGAAGATAGAAAACAAG 6330
Qy 781 CAATAACAGCATCATCTGTGCTGGAGTGACTGACCACACAGGCGCAAGACAGATTGG 840
Db 6331 CAATACACAGCTCTTC-----GTTTAAGAAATCTTGGTGGGAGATTACTGG 6378
Qy 841 AAACCCAAAAGCCAGGCTGAAAAAACCTGGAC-----CGCCTTGGGCTGCTTTTGC 894
Db 6379 GAACCCCTCCGTCGCCGCTCTGAATGCCAGGACGCTGTGAATGCTGGCAAGCCAAGCA 6438
Qy 895 ACTGATGAATACAGTGGTTACAAATAGATTGTAATAGGAAAAGAAAATACAGGCAATT 954
Db 6439 AACAAATAAGCAGTGGCTAGAAATTTGATCTACTCAAGATCAAGAAGATAACGGCAATT 6498
Qy 955 ATACCACTGGATCCACCATGCTGGAGCACAATTTACTATGTCTGCTACAGAACTCTG 1014
Db 6499 ATACACAGGGCTCCTGCTCTCTGAAATGTATGTAAGAGCTATACCATCCAC 6558
Qy 1015 TACAGTGATGATGGGAGAAATGAGTCTGTACAGAGAGCGCTGTGTGGAGCAAGATAAG 1074
Db 6559 TACAGTGAGCAGGAGTGGATGGAATGGAACCATACAGGCTGAAATCCTCCATGGTGGACAAG 6618
Qy 1075 ATATTTCAAGAAACAAGATTATACACAGGATGTCGTAATACTTTTCCACCAATT 1134
Db 6619 ATTTTGAAGGAAATACTAATACCAAGGACATGTGAAGAACTTTTCAACCCCAATC 6678
Qy 1135 ATTCACGTTTTATTAGAGTGAATCCTACCAATGGCAGCAGAAAATGGCCATGAAAATG 1194
Db 6679 ATTTCCAGGTTTATCCGTGTCATTCCTTAAACATGGAATCAAAGTATTACACTTCGCCCTG 6738
Qy 1195 GAGCTGCTCGGATGTCAGTTT 1215
Db 6739 GAACCTTTGGCTGTGATATT 6759

```

RESULT 5
 US-08-746-111-26
 Sequence 26, Application US/08746111
 Patent No. 6066778
 GENERAL INFORMATION:
 APPLICANT: Ginsburg, David
 APPLICANT: Cui, Jisong

Db 1326 GATTGGAAGCCAGAGTACATATAAAATCTTACAAAATTTGCTACAGCAATGACGGGAAGAC 1385
QY 1035 ATGAGTGTGTACAGAGAGCGCTGTGTGGAGCAAGATAAGATATTTTCAAGGAAACAAAGA 1094
Db 1386 CTGGCAATGTACAAAGTAAAGGACCAATGAAGAGATGCTCTTCTGCGAAATGTTGA 1445
QY 1095 TTATCACCAGATGTGCGTAATAACTTTTGGCCACCACCAATATTGCGACGTTTTATTAGAGT 1154
Db 1446 TAACAACACACCATATGCTAATCTTTTACACACCCCAATCAAGCTCAGTATGTAAGACT 1505
QY 1155 GAATCTTACCAATGGCAGCAGAAAATTTGCCATGAAAATGGAGCTGCTCGAGTGTCAAGT 1214
Db 1506 CTACCCCAAAATTTGTGAAGGCAATGCTACTTTAAGAATGAACTTCTTGCGCTGAGCT 1565

RESULT 7

US-08-659-235C-9
; Sequence 9, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quettermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,235C
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0034-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 619..2058
US-08-659-235C-9

Query Match 3.9%; Score 86.4; DB 2; Length 2303;
Best Local Similarity 50.2%; Pred. No. 3.3e-17;
Matches 241; Conservative 0; Mismatches 236; Indels 3; Gaps 1;

QY 735 ATGTTATGGAACACCTGGGAGTGGTGTGATCGCGATCCTCAATAACAGCATC 794
Db 1089 ATGCTCTGGGACCTGGGAATCGAAGTGGGATCATATCTATCAGCAATCAGAGCTTC 1148
QY 795 ATCTGTGTGAGTGTGAGTGTGACCAACAGGCGAGCAACAGTGTGGAACCCAAAGC 854

Db 1149 ATTAATC---ACCGAGCTCTTTTGGACTCCAGAAAGTGGTATCCCTACTATGCTCGACT 1205
QY 855 CAGGCTGAAAAAACCTGGACCGCTTGGGCTGCTTTTGGCCACTGATGAATACCAGTGGTT 914
Db 1206 TAATAAGAGGGCTTTATAATGCTTGGACAGCTGCTGAAATGACAGATGGCCATGAT 1265
QY 915 ACAATAGATTGAATAAGGAAAGAAAATAACAGGCAATTATACCACTGGATCCCAT 974
Db 1266 TCAGATAAATTTGCAAGAAAAATGAGAGTCACTGGTGTGTTATTACCAAGAGCAAAAG 1325
QY 975 GGTGGAGCACAATTAATCTATGCTGCTGCTACAGAACTCTTACAGTGTGATGGCAGAA 1034
Db 1326 GATTGGAAGCCAGAGTACATAAAATCTTACAAAATTTGCCCTACAGCAATGACGGGAAGC 1385
QY 1035 ATGACTGTGTACAGAGAGCGCTGTGTGGAGCAAGATAAGATATTTCAAGGAAACAAAGA 1094
Db 1386 CTGGCAATGTACAAAGTAAAGGCACCANTGAAGATGGTCTTCTGCGAAATGTTGA 1445
QY 1095 TTATCACCAGATGTGCGTAATAACTTTTGGCCACCACCAATATTGCGACGTTTTATTAGAGT 1154
Db 1446 TAACAACACACCATATGCTAATCTTTTACACACCCCAATCAAGCTCAGTATGTAAGACT 1505
QY 1155 GAATCTTACCAATGGCAGCAGAAAATTTGCCATGAAAATGGAGCTGCTCGAGTGTCAAGT 1214
Db 1506 CTACCCCAAAATTTGTGGAAGGCAATGCTACTTTAAGAATGAACTTCTTGCGCTGAGCT 1565

RESULT 8

US-08-480-229C-19
; Sequence 19, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quettermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0036-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-480-229C-19

Query Match

3.6%; Score 78.6; DB 2; Length 318;

[illegible]

RESULT 9
US-08-659-235C-19
; Sequence 19, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Bridg
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL

[illegible]

RESULT 10
US-08-480-229C-11
; Sequence 11, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0026-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:

TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 943-1931
TELEFAX: (510) 943-1189
TELEX: N.A.
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1384 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-07-607-538C-1

Query Match 3.5%; Score 76.8; DB 1; Length 1384;
Best Local Similarity 51.5%; Pred. No. 2.9e-14;
Matches 177; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
QY 878 CTGGGCTGCTTTTGGCCACTGATGATACCACTGGTTCACAAATAGATTGTAATAGGAAA 937
DB 323 CTGGGTTGGGGGAGCTACGTAACGATCAGTGGCTGCAGGTGGACCTGGGCTCCTCGA 382
QY 938 AGAAATAACAGGCATTATAACCACTGGATCCACCATGGTGAGGACACAATTAATATGTGT 997
DB 383 AGAGGTGACAGCATCATCACCAGGGGCCCGTAACCTTGGCTCTGTCCAGTTGTGG 442
QY 998 CTGCCCTACAGATCCTGTACAGTATGATGGCGAGAAATGGACTGTGTACAGAGAGCCTG 1057
DB 443 CATCCTACAGGTTGCTTACAGTAATGACAGTGCAGAACTGGACTGAGTACCAGGCCCA 502
QY 1058 GTGTGGAGCAAGATAGATATTTCAGGAAACAAGATTATCACCAGGATGTGCGTAATA 1117
DB 503 GCAGTGGCAGCAGTAGATCTTCCCTGGCACTGGGACAAACCACTCCCAAGAAGAACT 562
QY 1118 ACTTTTGGCCACCAATTAATGCACTGTTTATTAGAGTGAATCCTACCAATGGCAGCAGA 1177
DB 563 TGTTTGAGACGCCCATCTCGCTGCTATGTGCGCATCTGCTGTAGCTGGCCACACC 622
QY 1178 AAATGGCCATGAATGGAGCTGCTCGGATGTCAGTTTATTCCT 1221
DB 623 GCATCGCCCTGCGCTGGAGCTGCTGGGCTGTATTAGTGGCCACCT 666

RESULT 13
US-08-162-402B-1
Sequence 1, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LARocca, DAVID J.
TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162.402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1384 base pairs
TYPE: nucleic acid
STRADEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-162-402B-1

Query Match 3.5%; Score 76.8; DB 2; Length 1384;
Best Local Similarity 51.5%; Pred. No. 2.9e-14;
Matches 177; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
QY 878 CTGGGCTGCTTTTGGCCACTGATGATACCACTGGTTCACAAATAGATTGTAATAGGAAA 937
DB 323 CTGGGTTGGGGGAGCTACGTAACGATCAGTGGCTGCAGGTGGACCTGGGCTCCTCGA 382
QY 938 AGAAATAACAGGCATTATAACCACTGGATCCACCATGGTGAGGACACAATTAATATGTGT 997
DB 383 AGAGGTGACAGCATCATCACCAGGGGCCCGTAACCTTGGCTCTGTCCAGTTGTGG 442
QY 998 CTGCCCTACAGATCCTGTACAGTATGATGGCGAGAAATGGACTGTGTACAGAGAGCCTG 1057
DB 443 CATCCTACAGGTTGCTTACAGTAATGACAGTGCAGAACTGGACTGAGTACCAGGCCCA 502
QY 1058 GTGTGGAGCAAGATAGATATTTCAGGAAACAAGATTATCACCAGGATGTGCGTAATA 1117
DB 503 GCAGTGGCAGCAGTAGATCTTCCCTGGCACTGGGACAAACCACTCCCAAGAAGAACT 562
QY 1118 ACTTTTGGCCACCAATTAATGCACTGTTTATTAGAGTGAATCCTACCAATGGCAGCAGA 1177
DB 563 TGTTTGAGACGCCCATCTCGCTGCTATGTGCGCATCTGCTGTAGCTGGCCACACC 622
QY 1178 AAATGGCCATGAATGGAGCTGCTCGGATGTCAGTTTATTCCT 1221
DB 623 GCATCGCCCTGCGCTGGAGCTGCTGGGCTGTATTAGTGGCCACCT 666

RESULT 14
US-08-162-402B-7
Sequence 7, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LARocca, DAVID J.
TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162.402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER: 30,930
; FILING DATE: 25-MAR-1994
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: P66 38215
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1934 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-162-402B-7

Query Match      3.5%; Score 76.8; DB 2; Length 1934;
Best Local Similarity 51.5%; Pred. No. 3.6e-14;
Matches 177; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy 878 CTTGGGCTGCTTTGGCCACTGATCAATACCACTGGTTCACAAATAGATTGAATAGGAAA 937
Db 890 CTTGGGCTGCTTTGGCCACTGATCAATACCACTGGTTCACAAATAGATTGAATAGGAAA 937
Qy 938 AGAAATAACAGCATTAATACCACTGGTTCACAAATAGATTGAATAGGAAA 937
Db 950 AGGAGTGACAGGATCATCACCGGGGGCCGCTTGGCTGCTCCAGTTGGG 1009
Qy 998 CTGCTACAGAAATCCTGTACAGTATGATGGGCAAGAAATGACATGTGTACAGAGGCGTG 1057
Db 1010 CATCTACAAGTTGGCTTACAGTATGATGGGCAAGAAATGACATGTGTACAGAGGCGTG 1069
Qy 1058 GTGTGGAGCAGATTAAGATTTTCAAGGAACAAGATTAATCACCAGGATGTGCGTATA 1117
Db 1070 GGACTGGCAGCAGTAAGATCTTCCCTGGCAACTGGGCAACCACTCCCAAGAGAACT 1129
Qy 1118 ACTTTTGGCCACCAATATTGCAGCTTTTATTAGAGTGAATCCTACCAATGGCAGCAGA 1177
Db 1130 TGTGTGAGACCCCATCTGCTGCTATGTGCGCATCTCGCTGTAGCTGGCACAACC 1189
Qy 1178 AAATTGCCATGAAATGAGCTGCTGGATGTCAGTTTATTCCT 1221
Db 1190 GCATGCCCTCGCCTGGAGCTGCTGGGCTGTGTAGTGGCCACCT 1233

RESULT 15
US-08-484-891-7
; Sequence 7, Application US/08484891
; Patent No. 5935935
; GENERAL INFORMATION:
; APPLICANT: Connelly, Sheila
; APPLICANT: Kaleko, Michael
; APPLICANT: Smith, Theodore
; TITLE OF INVENTION: Adenoviral Vectors for
; TITLE OF INVENTION: Treatment of Hemophilia
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/484,891
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/218,335
; FILING DATE: 25-MAR-1994
; APPLICATION NUMBER: 08/074,920
; FILING DATE: 10-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 271010-273
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4629 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna primer
; FEATURE:
; NAME/KEY: Factor VIII cdna with
; NAME/KEY: B domain deleted
US-08-484-891-7

Query Match      2.9%; Score 63.4; DB 2; Length 4629;
Best Local Similarity 51.6%; Pred. No. 1.4e-09;
Matches 173; Conservative 0; Mismatches 156; Indels 6; Gaps 1;

Qy 878 CTTGGGCTGCTTTGGCCACTGATCAATACCACTGGTTCACAAATAGATTGAATAGGAAA 937
Db 4070 CTTGGGCTGCTTTGGCCACTGATCAATACCACTGGTTCACAAATAGATTGAATAGGAAA 4129
Qy 938 AGAAATAACAGCATTAATACCACTGGTTCACAAATAGATTGAATAGGAAA 937
Db 4130 TGAAGTCAACAGGAGTAATCACTCAGGAGTAAATCTCTCTTACCAGCATGTATGTGA 4189
Qy 998 CTGCTACAGAAATCCTGTACAGTATGATGGGCAAGAAATGAGCTGTGTACAGAGCCTG 1057
Db 4190 AGGAGTTCTCTATCTCCAGCAGTCAAGATGGCCATCAGTGGGACTCTCTTTTTCAGAATG 4249
Qy 1058 GTGTGGAGCAGATTAAGATTTTCAAGGAACAAGATTAATCACCAGGATGTGCGTATA 1117
Db 4250 GCA-----AAGTAAAGGTTTTTCAGGGAATCAAGACTCTCTTCCACACTCTGTGTGA 4303
Qy 1118 ACTTTTGGCCACCAATATTGCAGCTTTTATTAGAGTGAATCCTACCAATGGCAGCAGA 1177
Db 4304 CTCTAGACCCACCGTTACTGCTACCTTCCAAATTCACCCCAAGTTGGGTGCACC 4363
Qy 1178 AAATTGCCATGAAATGAGCTGCTCGGATGTGAG 1212
Db 4364 AGATTGCCCTGAGGATGGAGGTTCTGGGCTCGAG 4398

Search completed: January 21, 2003, 09:27:16
Job time : 148.864 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 00:35:10 ; Search time 446.376 Seconds
(without alignments)
11048.706 Million cell updates/sec

Title: US-10-060-830-2
Perfect score: 2190
Sequence: 1 atgcctctgtctctctgtct.....tttttaagaatccttga 2190

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	2026.4	92.5		2046	21	AAZ51872	Human Factor 8 Hom
2	1579.2	72.1		2428	22	AAH34855	Human colon cancer
3	1451.8	66.3		1907	24	ABK35994	Human neuropilin-H
4	1451.8	66.3		1907	24	ABK49567	Human cDNA encoding
5	1451.8	66.3		1907	24	ABN59799	Novel human coding
6	1158.4	52.9		1388	24	ABK83640	Human cDNA differe
7	594	27.1		594	23	AAK55485	DNA encoding novel
c	8	544.6	24.9	640	21	AAZ80396	Human colon cancer
	9	473.8	21.6	580	22	ABA60693	Human foetal liver

10	473.8	21.6	580	22	AAK08975	Human brain expres
11	473.8	21.6	580	22	AAK34866	Human bone marrow
12	473.8	21.6	580	22	AAI40582	Probe #9268 used t
13	467	21.3	467	22	ABA76819	Human foetal liver
14	467	21.3	467	22	AAK25439	Human brain expres
15	467	21.3	467	22	AAK51451	Human bone marrow
16	467	21.3	467	22	AAI57541	Probe #26227 used
17	232.4	10.6	588	22	ABA60998	Human foetal liver
18	232.4	10.6	588	22	AAK09290	Human brain expres
19	232.4	10.6	568	22	AAK35179	Human bone marrow
20	232.4	10.6	568	22	AAI40895	Probe #9581 used t
21	232.4	10.6	568	24	ABN09615	Human genome-deriv
22	228	10.4	228	22	ABA73598	Human foetal liver
23	228	10.4	228	22	AAK22044	Human brain expres
24	228	10.4	228	22	AAK48207	Human bone marrow
25	228	10.4	228	22	AAI54037	Probe #22723 used
26	228	10.4	228	24	ABN22095	Human genome-deriv
27	224.2	10.2	1620	22	AAK00615	Novel human protei
28	224.2	10.2	1761	22	AAK00614	Novel human protei
29	224.2	10.2	1768	22	AAK00613	Novel human protei
30	224.2	10.2	3594	22	AAK00660	Human TANGO 229 cD
31	202.8	9.3	412	20	AAK09008	Differentiation In
32	148.2	6.8	1871	21	AAK96736	Polyucleotide iso
C 33	131.2	6.0	324	24	AAK35993	Human expressed se
C 34	131.2	6.0	324	24	ABK49566	Human Neuropilin-H
35	91.6	4.2	6585	21	AAK60446	Murine factor v en
36	89.8	4.1	2209	22	AAH57553	Human brain cell s
C 37	89.8	4.1	4545	22	AAI26692	Human breast cance
C 38	89.8	4.1	4545	22	AAI26702	Human breast cance
39	89.8	4.1	4681	24	ABK86006	Human cDNA encodin
40	88.2	4.0	6893	20	AAK32182	Human coagulation
41	86.6	4.0	6909	16	AAK03920	Human Factor-V wll
42	86.6	4.0	6909	21	AAK60448	Human factor v cDN
43	86.6	4.0	6909	21	AAK39631	Human vth aggregat
44	86.6	4.0	6909	22	AAK07368	Human DNA encoding
45	86.6	4.0	6909	24	ABN95777	Gene #2275 used to

ALIGNMENTS

RESULT 1

AAZ51872
ID AAZ51872 standard; cDNA; 2046 BP.
XX AC AAZ51872;
XX DT 04-JUL-2000 (first entry)
XX DE Human Factor 8 Homologue cDNA.
XX DE Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
XX KW cerebroprotective; therapeutic; coagulation related disorder;
XX KW haemophilia; stroke; screening; ss.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX CDS .19..2028
XX FT FT /*tag= a
XX FT FT /product= "Factor 8 Homologue"
XX PN WO200012532-A1.
XX PD 09-MAR-2000.
XX PF 20-AUG-1999; 99WO-US19047.
XX PR 31-AUG-1998; 98US-0098521.
XX PA (ELIL) LILLY & CO ELI.
XX PI Rosteck PRJ, Su W, Li XM;

XX
DR
DR
DR
PT
PT
XX
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
SQ

WPI; 2000-256580/22.
P-PSDB; AAY70539.
Factor 8 homolog polypeptides and nucleic acids encoding them for treating coagulation related disorders such as hemophilia and stroke
Claim 1: Page 61-64; 68pp; English.
The present sequence is a cDNA encoding human Factor 8 homologue (F8H) which is a coagulation cofactor. The protein is selectively expressed in hematopoietic, heart and reproductive tissues. It has haemostatic and cerebroprotective activities. The F8H contains a Factor 5/8 signature and is useful as a therapeutic for treating coagulation related diseases such as haemophilia and stroke. The nucleic acid is useful as a hybridisation probe and amplification primer for detecting deficiencies in the level of F8H mRNA, for screening F8H gene mutations and for monitoring regulation of gene expression. Fragments of the nucleic acid are also useful as diagnostic probes and primers, and can be used in screening methods such as those using DNA chips. the present sequence is also useful as a target to screen therapeutically useful modulators of the F8H.

Sequence 2046 BP; 615 A; 457 C; 467 G; 507 T; 0 other;

Query Match 92.5%; Score 2026.4; DB 21; Length 2046;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2027; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 163 TGGGAGATCCGTTGAAGATGGAGAGAGTTCGCATCAATTTGGTGACATTGACATT 222
DB 1 TGGGAGATCCGTTGAAGATGGAGAGAGTTCGCATCAATTTGGTGACATTGACATT 60
QY 223 GAAGATTCTGATCTTGTACATTTAATTTACTTTGAGAATTTAATGGAATGGAGTCAGC 282
DB 61 GAAGATTCTGATCTTGTACATTTAATTTACTTTGAGAATTTAATGGAATGGAGTCAGC 120
QY 283 AGAAGTGAATAGCAATACATCTGCTGGGTTGCAATGCAATTCATTTGAATCA 342
DB 121 AGAAGTGAATAGCAATACATCTGCTGGGTTGCAATGCAATTCATTTGAATCA 180
QY 343 AAAGCAATGAATCAGATCTGTTGATGAGTGAATCCATTTCTGGAGCGGATTT 402
DB 181 AAAGCAATGAATCAGATCTGTTGATGAGTGAATCCATTTCTGGAGCGGATTT 240
QY 403 TTGGCTCATCTCTGTTATAGATAAACAAGATCTAATTTACTTTTGGACACTGCATCC 462
DB 241 TTGGCTCATCTCTGTTATAGATAAACAAGATCTAATTTACTTTTGGACACTGCATCC 300
QY 463 AATTTTGGAACTGAGTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 522
DB 301 AATTTTGGAACTGAGTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 360
QY 523 GAGATATCTGGAACTTCTTCATGATAGATAGATATCTGCGCAATTTGATGAGTGGT 582
DB 361 GAGATATCTGGAACTTCTTCATGATAGATAGATATCTGCGCAATTTGATGAGTGGT 420
QY 583 GTGCATGAGGAGTGTGCAACACGTTGGCGGCCCAATTCAGTGTGTTAATAGTAAA 642
DB 421 GTGCATGAGGAGTGTGCAACACGTTGGCGGCCCAATTCAGTGTGTTAATAGTAAA 480
QY 643 GGTATTCCTATTATGAAGTTCTTTTGGCTAACACGTCACATCTGTGGTGGGACATTA 702
DB 481 GGTATTCCTATTATGAAGTTCTTTTGGCTAACACGTCACATCTGTGGTGGGACATTA 540
QY 703 TCTAAGTCTTTTACATTTAAGACAGTGGATGTTATGGAACACTGGGATGGAGTCT 762
DB 541 TCTAAGTCTTTTACATTTAAGACAGTGGATGTTATGGAACACTGGGATGGAGTCT 600
QY 763 GTGTGATCGCGGATCTCAATTAACAGATCATCTGTGCTGGAGTGACTGACCACACA 822
DB 601 GTGTGATCGCGGATCTCAATTAACAGATCATCTGTGCTGGAGTGACTGACCACACA 660

QY 823 GGGCAAGAGAACAGTTGGAAACCCCAAAAGCCAGGCTGAAAGAAACCTGGACCGCTTGG 882
DB 661 GGGCAAGAGAACAGTTGGAAACCCCAAAAGCCAGGCTGAAAGAAACCTGGACCGCTTGG 720
QY 883 GCTGCTTTTGGCACTGATGAATACCAAGTGGTTTACAAATAGATTTGAATAGGAAAGAAA 942
DB 721 GCTGCTTTTGGCACTGATGAATACCAAGTGGTTTACAAATAGATTTGAATAGGAAAGAAA 780
QY 943 ATACAGAGCAATATACCACTGGATCCACCATGGTGGAGCACAATTACTATGTCTGCC 1002
DB 781 ATACAGAGCAATATACCACTGGATCCACCATGGTGGAGCACAATTACTATGTCTGCC 840
QY 1003 TACAGATCTCTGTACAGTGTATGGGAGAAATGGATGTGTACAGAGAGCTGGTGTG 1062
DB 841 TACAGATCTCTGTACAGTGTATGGGAGAAATGGATGTGTACAGAGAGCTGGTGTG 900
QY 1063 GAGCAAGATAAGATATTTCAAGGAACAAAGATTTACACAGATGTGGTGAATAACTTT 1122
DB 901 GAGCAGATAGATATTTCAAGGAACAAAGATTTACACAGATGTGGTGAATAACTTT 960
QY 1123 TTGCCACCAATTTATGACAGTGTATTTATAGAGTGAATCTTACCCTATGGCAGAGAAAT 1182
DB 961 TTGCCACCAATTTATGACAGTGTATTTATAGAGTGAATCTTACCCTATGGCAGAGAAAT 1020
QY 1183 GCCATGAATTTGGAGCTGCTCGGATGTCAGTGTATTTCTTAAAGTGTCTCTCCAAACTT 1242
DB 1021 GCCATGAATTTGGAGCTGCTCGGATGTCAGTGTATTTCTTAAAGTGTCTCTCCAAACTT 1080
QY 1243 ACTCAACTCCACCTCTCGGAACAGCAATGACCTCAAAACACACTACAGCCCTCCAAA 1302
DB 1081 ACTCAACTCCACCTCTCGGAACAGCAATGACCTCAAAACACACTACAGCCCTCCAAA 1140
QY 1303 ATAGCAAGGTCTGTCGCCCAAAATTTACGCAACCACTACAACTTCGAGTAGCAATGAA 1362
DB 1141 ATAGCAAGGTCTGTCGCCCAAAATTTACGCAACCACTACAACTTCGAGTAGCAATGAA 1200
QY 1363 TTTCTGTGACAGACAGAACAACTGCTGAGTCTCTGATATCAGAAATACCTACCTTAAT 1422
DB 1201 TTTCTGTGACAGACAGAACAACTGCTGAGTCTCTGATATCAGAAATACCTACCTTAAT 1260
QY 1423 CCAATGTAAACCAAGATGTAGCGCTGCTGCAGTCTTGTCCCTGCTGCTGCTGCTGCT 1482
DB 1261 CCAATGTAAACCAAGATGTAGCGCTGCTGCAGTCTTGTCCCTGCTGCTGCTGCTGCT 1320
QY 1483 CTCCTACTCTCATTTCTCATATTTAGTGTGCTTGGCACTGGAGAAACAGAAAGAAAAA 1542
DB 1321 CTCCTACTCTCATTTCTCATATTTAGTGTGCTTGGCACTGGAGAAACAGAAAGAAAAA 1380
QY 1543 ACTGAAGCACTATGACTTACCTTTACTGGGACCGGAGGTTGTTGGAAGGAATGAAG 1602
DB 1381 ACTGAAGCACTATGACTTACCTTTACTGGGACCGGAGGTTGTTGGAAGGAATGAAG 1440
QY 1603 CAGTTTCTCTGCAAAAGCAGTGGACCATGAGGAAACCCCAAGTTGCTATAGCAGCAGC 1662
DB 1441 CAGTTTCTCTGCAAAAGCAGTGGACCATGAGGAAACCCCAAGTTGCTATAGCAGCAGC 1500
QY 1663 GAAGTTAATCACTGAGTCCCAAGAGAGTACCACAGTGTCTGAGGCTGACTCTCAGAG 1722
DB 1501 GAAGTTAATCACTGAGTCCCAAGAGAGTACCACAGTGTCTGAGGCTGACTCTCAGAG 1560
QY 1723 TATGCTCAGCCACTGGTAGGAGGAATTTGTTACACTTTCATCAAGATCTACCTTTAAA 1782
DB 1561 TATGCTCAGCCACTGGTAGGAGGAATTTGTTACACTTTCATCAAGATCTACCTTTAAA 1620
QY 1783 CCAGAAGAGAAAGAGAGAGGCTATGCGACCTAGATTCCTTACAACTACCAGGGGAG 1842
DB 1621 CCAGAAGAGAAAGAGAGAGGCTATGCGACCTAGATTCCTTACAACTACCAGGGGAG 1680
QY 1843 GAAGTTTATCATGCTGATGCTGAACCTCCCAATTTACGGGCTGAGTATGCAACCCCA 1902
DB 1681 GAAGTTTATCATGCTGATGCTGAACCTCCCAATTTACGGGCTGAGTATGCAACCCCA 1740
QY 1903 ATCATGTGACATGTCTAGGGCAGCCCAACACTTCAGTTGTTGCTCAGCCCTCCACATCCACT 1962

Db	161	CAGCAAGGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTTACATCC	220
Qy	121	ATAAACTACCACAGACCTATCCCAACAGCACTGTTTGTGAATGGGAGATCCGTTGAAG	180
Db	221	ATAAACTACCACAGACCTATCCCAACAGCACTGTTTGTGAATGGGAGATCCGTTGAAG	280
Qy	181	ATGGGAGAGAGATTTCGCATCAAAATTGGTGACTTTTGACATTGAAGATTCTGATCTTGT	240
Db	281	ATGGGAGAGAGATTTCGCATCAAAATTGGTGACTTTTGACATTGAAGATTCTGATCTTGT	340
Qy	241	CACCTTAATTACTGTGAGATTTATATGGAATTGGAGTCAGCAGCACTGAAATAGGCAAA	300
Db	341	CACCTTAATTACTGTGAGATTTATATGGAATTGGAGTCAGCAGCACTGAAATAGGCAAA	400
Qy	301	TACTGTGGTCTGGGGTGCAAAATGAACCAATCAATGAATCAAAAGGCAATGAATCACA	360
Db	401	TACTGTGGTCTGGGGTGCAAAATGAACCAATCAATGAATCAAAAGGCAATGAATCACA	460
Qy	361	TTGCTGTTCAATGAGTGGAAATCCATGTTCTGGACGGGATTTTGGCCCTCAATCTTGTT	420
Db	461	TTGCTGTTCAATGAGTGGAAATCCATGTTCTGGACGGGATTTTGGCCCTCAATCTTGTT	520
Qy	421	ATAGATAACAAGATCAATTACTTGTGTGGACACTGCATCCAATTTTTCGAACCTTGAG	480
Db	521	ATAGATAACAAGATCAATTACTTGTGTGGACACTGCATCCAATTTTTCGAACCTTGAG	580
Qy	481	TTCAGTAAAGTACTGCCAGCTGGTGTCTGCTTTCCTTTTGTGAGATATCTGGAAACAAT	540
Db	581	TTCAGTAAAGTACTGCCAGCTGGTGTCTGCTTTCCTTTTGTGAGATATCTGGAAACAAT	640
Qy	541	CCTCATGGATATAGAGATTCTTCGCCATTTGTCGATGGCTGGTGTCATGCGAGGATGAG	600
Db	641	CCTCATGGATATAGAGATTCTTCGCCATTTGTCGATGGCTGGTGTCATGCGAGGATGAG	700
Qy	601	TCAAAACAGTGTGGCGGCCAAATCAGTGTGTAATTAGTAAAGGTATTCCTATTATGAA	660
Db	701	TCAAAACAGTGTGGCGGCCAAATCAGTGTGTAATTAGTAAAGGTATTCCTATTATGAA	760
Qy	661	AGTTCTTTGGCTTAACAACTGCATCTGTGTGGGACACTTATCTACAAAGTCTTTTTACA	720
Db	761	AGTTCTTTGGCTTAACAACTGCATCTGTGTGGGACACTTATCTACAAAGTCTTTTTACA	820
Qy	721	TTTAAGACAAGTGGATGTTATGGAACATGCGGGATGGAGTCTGGT- GTGATCGCGGATCC	779
Db	821	TTTAAGACAAGTGGATGTTATGGAACATGCGGGATGGAGTCTGGTGTGATCGCGGATCC	880
Qy	780	TCAATAACAGCATCATCTGCTGGAGTGGACTGACCACACAGGGCAAGAGACAGTTG	839
Db	881	TCAATAACAGCATCA- CTGTGCTGGAGTGGACTGACCACACAGGGCAAGAGACAGTTG	939
Qy	840	GAAACCC- AAAAAAGCAGGCTGAAAAACCTGGACCGCCTTGGGCTGCTTTTGGCACTG	898
Db	940	GAAACCCAAAAAAGCCAGGCTGAAAAACCTGGACCGCCTTGGG- TCGTTTTGCCACTG	998
Qy	899	ATGAATACCAGTGGTTTACAAATAGATTGGAATAAGGAAAAAGAAAATAACAGGCATTATA	958
Db	999	ATGAATACCAGTGGTTTACAAATAGATTGGAATAAGGAAAAAGAAAATAACAGGCATTATA	1058
Qy	959	CCACTGGATCCACCATGGTGGAGCACAATTAATACTATGCTGCTCGCTACAGATCTCTGATA	1018
Db	1059	CCACTGGATCCACCATGGT- GAGACACAATTAATACTATGCTGCTCGCTACAGATCTCTGATA	1117
Qy	1019	GTGATGATGGCAGAAATGGACTGTGTACAGAGAGCCCTGGT- GTGGAGCAAGATAAGATA	1077
Db	1118	GTGATGATGGCAGAAATGGACTGTGTACAGAGAGCCCTGGTGGTGGACAGATAGATA	1177
Qy	1078	TTTCAAGGAACAAA- - GATTATCACCGGAT- GTGCGCTAATAACTTTTTTCCCAACCAATT	1134
Db	1178	TTTCAAGGAACAAAAGAAATTAACACGATGGTGGCTAATAACTTTTTTCCCAACCAATT	1237
Qy	1135	ATTCCAGCTTTTATTAGATGGAATCTCCCAATGGCAGCAGAAAAATTTGCCATGAAATG	1194
Db	1238	ATTCCAGC- TTTATTAGATGGAATCTCCCAATGGCAGCAGAAAAATTTGCCATGAAATG	1296

Qy	1195	GAGTGTCTCGGATGTCAAGTTTATTCTTAAGTGTCCTCCAAAACCTTACTCAACCTCCA	1254
Db	1297	GAGTGTCTCGGATGTCAAGTTTATTCTTAAGGTGCTCTCCAAAACCTTACTCAACCTCCA	1356
Qy	1255	CCTCCTCGGAACAGCAATGACCTCAAAAAACACTACAGCCCCTCCAAAAA---TAGGCCAAA	1311
Db	1357	CCTCCTCGGAACAGCAATGACCTC-AAAAACACTACAGCCCCTCCAAAAAAATTTAGCCAAA	1415
Qy	1312	GGTGTGCCCCCAAATTTTAGCAACAACACTACAACCTCGCAGTAGCAATTTCTCTGCA	1371
Db	1416	GGTGTGCCCCCAAATTTTAGCAACAACACTACAACCTCGCAGTAGCAATTTCTCTGCA	1475
Qy	1372	CAGACAGAACAACAACCTGCAGTCTCTGATATCAGAAATACTACCGTACTCCAAATGTA	1431
Db	1476	CAGACAGAACAACAACCTGCAGTCTCTGATATCAGAAATACTACCGTACTCCAAATGTA	1535
Qy	1432	ACCAAGATGTAGCGCTGGCTGCAGTTCTTGTCCTGTGCTGGTCATGGTCTCACTACT	1491
Db	1536	ACCAAGATGTAGCGCTGGCTGCAGTTCTTGTCCTGTGCTGGTCATGGTCTCACTACT	1595
Qy	1492	CTCATTCATATATTAGTGTGTGCTGGCACTGGAGAAACAGAGAAAAAACTGAAGGC	1551
Db	1596	CTCATTCATATATTAGTGTGTGCTGGCACTGGAGAAACAGAGAAAAAACTGAAGGC	1655
Qy	1552	ACCTATGACTTACTTACTGGGACCGGGCAGGT	1584
Db	1656	ACCTATGACTTACTTACTGGGACCGGGCAGGT	1688
 RESULT 4 ABK49567			
ID	ABK49567 standard; cDNA; 1907 bp.		
XX	ABK49567;		
XX			
DT	15-JUL-2002 (first entry)		
DE	Human cDNA encoding Neuropilin-Hy2.		
XX			
KW	Human; ss; gene; neuropilin-Hy2; chromosome 6q21; neuronal growth;		
KW	nerve regeneration; neurodegenerative disease; learning disorder;		
KW	memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;		
KW	organ growth; nervous system lesion; cancer; cell proliferation;		
KW	cell differentiation; stem cell growth factor activity;		
KW	Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis;		
KW	haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;		
KW	platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;		
KW	refeupusion; food supplement; DNA microarray.		

RESULT 4

RESULT 4
ABK49567

ABK43567
ID ABK49567 standard: cDNA: 1907 BP.

XX
ID ABN43367

AC ABK49567;

XX
XX

DT 15-JUL-2002 (first entry)

Human cDNA encoding Neurotrophin-Hy2.

Human; ss: gene; neurotrophin-Hy2; chromosome 6q21; neuronal growth; nerve regeneration; neurodegenerative disease; learning disorder; memory disorder; Alzheimer's disease; angiogenesis; neovascularisation; organ growth; nervous system lesion; cancer; cell proliferation; cell differentiation; stem cell growth factor activity; parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis; haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia; platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis; reperfusion; food supplement. cDNA microarray.

XX	Homo sapiens.
OS	
XX	
PH	Key
XX	CDS
FT	Location/Qualifiers
FT	101..1258
FT	/tag= a
FT	/product= "Neuropilin-Hy2"
FT	/note= "This coding sequence is specifically claimed in
FT	claim 1"
FT	101..160
FT	/tag= b
FT	161..1255
FT	/tag= c
FT	/label= Mature_Neuropilin_Hy2
PN	
XX	WO200222780-A2.
XX	
PD	21-MAR-2002.
XX	
PF	11-SEP-2001; 2001WO-US28590.
XX	
PR	11-SEP-2000; 2000US-0659671.
PR	06-SEP-2001; 2001US-0659671.
XX	

QY 1135 ATTGCACGTTTATTAGAGTGAATCCTACCCAAATGGCAGCAAAATTCGCATGAAATG 1194
DB 1238 ATTGCACG-TTATTAGAGTGAATCCTACCCAAATGGCAGCAAAATTCGCATGAAATG 1296
QY 1195 GAGCTGCTGGATGTCAGTTTATTCTTAAAGTCTGCTCCCAAAATCTTACTCAACCTCCA 1254
DB 1297 GAGCTGCTGGATGTCAGTTTATTCTTAAAGTCTGCTCCCAAAATCTTACTCAACCTCCA 1356
QY 1255 CTTCTCTCGAAGCAGCAATGACCTCAAAACACTAGAGCCCTCCCAAAA---TAGCCAAA 1311
DB 1357 CTTCTCTCGAAGCAGCAATGACCTC-AAACACTAGAGCCCTCCCAAAAATTTAGCCAAA 1415
QY 1312 GTCGTGCCCCAAAATTTACGCAACCTACAACTCGCAGTAGCAATGAATTTCTCTGCA 1371
DB 1416 GTCGTGCCCCAAAATTTACGCAACCTACAACTCGCAGTAGCAATGAATTTCTCTGCA 1475
QY 1372 CAGACAGAACAAACACTGCCAGTCTGATATCAGAAATACTACCGTAACCTCAAAATGTA 1431
DB 1476 CAGACAGAACAAACACTGCCAGTCTGATATCAGAAATACTACCGTAACCTCAAAATGTA 1535
QY 1432 ACCAAGATGAGCGCTGGCTGCACTTCTTGTCCCTGTGCTGGTCATGGTCTCACTACT 1491
DB 1536 ACCAAGATGAGCGCTGGCTGCACTTCTTGTCCCTGTGCTGGTCATGGTCTCACTACT 1595
QY 1492 CTCATTCTCATATTAGTGTGCTGGCACTGGAGAAACAGAAAGAAAACCTGAAGGC 1551
DB 1596 CTCATTCTCATATTAGTGTGCTGGCACTGGAGAAACAGAAAGAAAACCTGAAGGC 1655
QY 1552 ACTATGACTTACCTTACTGGACCGGCAGGT 1584
DB 1656 ACCATGACTTACCTTACTGGACCGGCAGGT 1688

RESULT 6

ID ABR83640 standard; cDNA; 1388 BP.

AC ABR83640;

DT 14-AUG-2002 (first entry)

DE Human cDNA differentially expressed in granulocytic cells #211.

KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.

XX WO200228999-A2.

PN 11-APR-2002.

XX 03-OCT-2001; 2001WO-US30821.

PF 03-OCT-2000; 2000US-237189P.

XX (GENE-) GENE LOGIC INC.

PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

XX WPI; 2002-435328/46.

DR Detecting granulocyte activation by detecting differential expression

PT of genes associated with granulocyte activation, which serves as

PT diagnostic markers that is useful for monitoring disease states and

PT drug toxicity .

XX PS Claim 1; SEQ ID No 211; 114pp; English.
XX CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1388 BP; 432 A; 289 C; 302 G; 364 T; 1 other;

Query Match 52.9%; Score 1158.4; DB 24; Length 1388;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1358; Conservative 0; Mismatches 12; Indels 23; Gaps 17;
QY 68 GTGATGATGGACACACTCTACTAGGCCCTGAGAGTGGAAACCTTACATCCATAAAT 127
DB 1 GTGATGATGGACACACTCTACTAGGCCCTGAGAGTGGAAACCTTACATCCATAAAT 60
QY 128 ACCCAGACCTATCCCAACAGCAGCTGTTTGTGATGGAGATCCGTGAAGATGGGAG 187
DB 61 ACCCAGACCTATCCCAACAGCAGCTGTTTGTGATGGAGATCCGTGAAGATGGGAG 120
QY 188 AGAGAGTTCGCATCAAAATTTGGTGAATTTGACATTTGAAGATTTCTGATCTTCTACTTTA 247
DB 121 AGAGAGTTCGCATCAAAATTTGGTGAATTTGACATTTGAAGATTTCTGATCTTCTACTTTA 180
QY 248 ATTACTTGAGAAATTTAATGAATTTGAGTGCAGCAGAACTGAATAGCAAAATCTGTG 307
DB 181 ATTACTTGAGAAATTTAATGAATTTGAGTGCAGCAGAACTGAATAGCAAAATCTGTG 240
QY 308 GTCCTGGGTTGCAAAATGAACCATTTCAATTAATCAAAAGGCAATCAATACATTCCTGT 367
DB 241 GTCCTGGGTTGCAAAATGAACCATTTCAATTAATCAAAAGGCAATCAATACATTCCTGT 300
QY 368 TCATGAGTGAATCCATGTTTCTGGACCGGATTTTGGCCTCATACATCTCTGTATAGATA 427
DB 301 TCATGAGTGAATCCATGTTTCTGGACCGGATTTTGGCCTCATACATCTCTGTATAGATA 360
QY 428 AAC-AAGATCTAATTAATCTGTTTGGACACTGCATCCAA-TTTTFTGGAACTGAGTTCCAG 485
DB 361 AACAAAGATCTAATTAATCTGTTTGGACACTGCATCCAAATTTTFTGGAACTGAGTTCCAG 420

```
Qy 486 TAAGTACTG-CCAGCTGGTGTCTGCTCTCTTTTGTGAGATATCTGGAACAATTCCTC 544
Db 421 TAAGTACTGCCAGCTGGTGTCTGCTCTCTTTTGTGAGATATCTGGAACAATTCCTC 480
Qy 545 ATGGATATAGAGATTCCTCGCCATGTGTCATGGTGTGTCATGTCAGGAGTAGTGTCAA 604
Dy 481 ATGGATATAGAGATTCCTCGCCATGTGTCATGGTGTGTCATGTCAGGAGTAGTGTCAA 540
Qy 605 ACAGTTGGGGCGGCAAAATCAGTGTGTAATAGTAAGGATTCCTTATTAATGAAGTT 664
Dy 541 ACAGTTGGGGCGGCAAAATCAGTGTGTAATAGTAAGGATTCCTTATTAATGAAGTT 600
Qy 665 CTTTGGCTAACAGCTCACATCTGCTGCTGGGACACTT--ATCTACAAGTC-TTTTACAT 721
Dy 601 CTTTGGCTAACAGCTCACATCTGCTGCTGGGACACTTGTACTACAAGTC-TTTTACAT 660
Qy 722 TTAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGT--GTGATCGCGATCCT 780
Dy 661 TTAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGGTGGTATCGCGATCCT 720
Qy 781 CAATAACAGCATCATCTGCTGGAGTGGACTGACCACACAGGCGCAAGAGAAGAGTTGG 840
Dy 721 CAATAACAGCATCA-CTGTGCTGGAGTGGACTGACCACACAGGCGCAAGAGAAGAGTTGG 779
Qy 841 AAACCC-AAACCCAGGCTGAAAACCTGGACCGCTTGGGCTGCTTTTGGCACTGA 899
Dy 780 AAACCCAAAAAAGCCAGGCTGAAAACCTGGACCGCTTGGG-TGCTTTTGGCACTGA 838
Qy 900 TGAATACAGGTTTACAAATAGATTTGAATAAGGAAAGAAATAACAGGCATTAATAC 959
Dy 839 TGAATACAGGTTTGAATAAGATTTGAATAAGGAAAGAAATAACAGGCATTAATAC 898
Qy 960 CACTGGATCCACCATGTGGAGGACAAATTAATGCTGCTGCTACAGAAATCCTGTACAG 1019
Dy 899 CACTGGATCCACCATGTG-GAGCAATTAATGCTGCTGCTACAGAAATCCTGTACAG 957
Qy 1020 TGATGATGGGAGAAATGGAGTGTACAGAGCCCTGGT-GTGAGCAAGATAGATAT 1078
Dy 958 TGATGATGGGAGAAATGGAGTGTACAGAGCCCTGGTGTGGAGCAAGATAGATAT 1017
Qy 1079 TTCAGGAAACAAA--GATTAATACCAGGAT-GTGGCTTAATAACTTTTGGCACCATA 1135
Dy 1018 TTCAGGAAACAAAAGAAATTAATCAGGAGTGGTGGCTTAATAACTTTTGGCACCATA 1077
Qy 1136 TTGACGTTTATAGAGTGAATCCTACCCAAATGGCAGCAAGAAATTCCTGAAATGG 1195
Dy 1078 TTGACG-TTTATAGAGTGAATCCTACCCAAATGGCAGCAAGAAATTCCTGAAATGG 1136
Qy 1196 AGCTGCTGGATGTCAGTTTATCCTAAAGTGTGCTTCCAAAATTTACTCAACCTCCAC 1255
Dy 1137 AGCTGCTGGATGTCAGTTTATCCTAAAGTGTGCTTCCAAAATTTACTCAACCTCCAC 1196
Qy 1256 CTCCTCGGAACAGCAATGACCTCAAAACACTACAGCCCTCCAAAAA---TAGCCAAG 1312
Dy 1197 CTCCTCGGAACAGCAATGACCTC-AAACACTACAGCCCTCCAAAAAATTTAGCCAAG 1255
Qy 1313 GTCGTGCCCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATGAATTCCTGCAC 1372
Dy 1256 GTCGTGCCCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATGAATTCCTGCAC 1315
Qy 1373 AGACAGAACCAACCACTGCCAGTCTGTATATCAGAAATACCTA---CCTAACTCCAAATG 1429
Dy 1316 AGACAGAACCAACCACTGCCAGTCTGTATATCAGAAATACCTA---CCTAACTCCAAATG 1375
Qy 1430 TAACCAAGATGT 1442
Dy 1376 TAACCAAGATGT 1388
RESULT 7
AAS65485
ID AAS65485 standard; cDNA; 594 BP.
```

```
XX AAS65485;
AC 13-FEB-2002 (first entry)
DT DNA encoding novel human diagnostic protein #1289.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
PF 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX P-PSDB; ABG01298.
DR New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID No 1289; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 594 BP; 175 A; 160 C; 139 G; 120 T; 0 other;
```

```
Query Match 27.1%; Score 594; DB 23; Length 594;
Best Local Similarity 100.0%; Pred. No. 1.3e-175;
Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1597 ATGAGCAGTTCCTCTGCAAAAGCAGTGGACCATGAGAAACCCCATGTCGTATAGC 1656
Dy 1 ATGAGCAGTTCCTCTGCAAAAGCAGTGGACCATGAGAAACCCCATGTCGTATAGC 60
Qy 1657 AGCAGCGAAGTAAATCACTGAGTCCCAAGAGAAGTCAACACAGTGTCTGAGGCTGACTCT 1716
Dy 61 AGCAGCGAAGTAAATCACTGAGTCCCAAGAGAAGTCAACACAGTGTCTGAGGCTGACTCT 120
Qy 1717 GCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGTTGTACACTTCATCAAGATCTACC 1776
Dy 1717 GCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGTTGTACACTTCATCAAGATCTACC 1776
```

Db 121 GCAGAGTATGTCAGGCACCTGGTAGGAGGAATTTGGTACACTTTCATCAAGATCTACC 180
QY 1777 TTTAAACCAAGAAAGAAAGCAGGCTATGCGAGACTAGATCCTTACAACTCACCA 1836
Db 181 TTTAAACCAAGAAAGAAAGCAGGCTATGCGAGACTAGATCCTTACAACTCACCA 240
QY 1837 GGCAGAGAAATTTATCATGCTATGCTGAACCACTCCCAATACGGGGCTGAGTATGCA 1896
Db 241 GGCAGAGAAATTTATCATGCTATGCTGAACCACTCCCAATACGGGGCTGAGTATGCA 300
QY 1897 ACCCAATCATCATGACATGTCAGGGCACCCACAACTTTCAGTTGGTCAGCCCTCCACA 1956
Db 301 ACCCAATCATCATGACATGTCAGGGCACCCACAACTTTCAGTTGGTCAGCCCTCCACA 360
QY 1957 TCCACTTTCAAGGCTACGGGGAACCAACTCCGCCACTAGTGGGAACCTTACAACTACACTT 2016
Db 361 TCCACTTTCAAGGCTACGGGGAACCAACTCCGCCACTAGTGGGAACCTTACAACTACACTT 420
QY 2017 CTCTCAGGACTGACAGCTGCTCTCAGCCAGGCCAGTATGATACCCCGAAAGCTGGG 2076
Db 421 CTCTCAGGACTGACAGCTGCTCTCAGCCAGGCCAGTATGATACCCCGAAAGCTGGG 480
QY 2077 AAGCCAGGCTACCTGCCAGACGAAATGGTGTACCAAGTCCAGAGACACAAAGAA 2136
Db 481 AAGCCAGGCTACCTGCCAGACGAAATGGTGTACCAAGTCCAGAGACACAAAGAA 540
QY 2137 GTATCAGGAGCAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGA 2190
Db 541 GTATCAGGAGCAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGA 594

RESULT 8
AAZ80396/c
ID AAZ80396 standard; cDNA; 640 BP.
XX
AC AAZ80396;
XX
DT 07-APR-2000 (first entry)
XX
Human colon cancer cell line SW480 cDNA clone SEQ ID NO:480.
DE
XX
Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
XX hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO964576-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-IB01062.
XX
PR 10-JUN-1998; 98US-0088801.
XX
PA (FARB) BAYER CORP.
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
DR WPI; 2000-087220/07.
XX
PT Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer
XX
PS Claim 15; Page 322; 469pp; English.
XX
CC AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which

CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g.
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC can be used for developing agents for the diagnosis and treatment of
CC disorders involving unwanted cell proliferation, such as neoplasia,
CC dysplasia or hyperplasia.
XX
SQ Sequence 640 BP; 148 A; 134 C; 160 G; 192 T; 6 other;
Query Match 24.9%; Score 544.6; DB 21; Length 640;
Best Local Similarity 96.7%; Pred No. 4.6e-160;
Matches 616; Conservative 0; Mismatches 15; Indels 6; Gaps 6;
QY 1121 TTTTGGCCACCAATTTATGACGTTTATTAGAGTGAATCTTACCCATGSCAGCAGAAA 1180
Db 632 TTTTGGCCACCAATTTATGACG-TTTATTAGAGT-AATCCTTACCCATGSCAGCAG-AAA 576
QY 1181 TTGGCCATGAATTTGACGTCGCGATGTCAGTTTATTCTTAAGTGTCTCTCCAAAAC 1240
Db 575 TTGGCCATGAATTTGACGTCGCGATGTCAGTTTATTCTNTAA-GTNGTCTTCCAAAAC 518
QY 1241 TTACTCAACCTCCACCTCTCGGACACGCAATGACCTCAAAACACTACAGCCCTCCAA 1300
Db 517 TTACTCAACCTCCACCTCTTNGACAGCAATGNCCTCAAAACACTACAGCCCTCC-A 459
QY 1301 AAATAGCCAAAGGTGTCGCCCAAAATTTACGCAACACTACAACTTCGCGAGTAGCAATG 1360
Db 458 AAATAGCCAAAGGTGTCGCCCAAAATTTACGCAACACTACAACTTCGCGAGTAGCAATG 399
QY 1361 AATTTCTTCGACAGACAGAAACAACTGCCAGTCTCTATATCAGAAATACTACCGTAA 1420
Db 398 AATTTCTTCGACAGACAGAAACAACTGCCAGTCTCTATATCAGAAATACTACCGTAA 339
QY 1421 CTCCAAATGTACCAAAGATGTAGCGCTGGCTGCAGTTCTTGTGCCCTGTGCTGGTCTATGG 1480
Db 338 CTCCAAATGTACCAAAGATGTAGCGCTGGCTGCAGTTCTTGTGCCCTGTGCTGGTCTATGG 279
QY 1481 TCCTCAGTACTCTCATTTCTCATATTAGTGTGTGGCACTGGAGAAACAGAAAGAAA 1540
Db 278 TCCTCAGTACTCTCATTTCTCATATTAGTGTGTGGCACTGGAGAAACAGAAAGAAA 219
QY 1541 AAACCTGAAGGACCTATGACCTTACTTGGACCGGCGAGTGGTGGAAAGAAATGA 1600
Db 218 AAACCTGAAGGACCTATGACCTTACTTGGGCGCGGAGGTTGGTGGAAAGAAATGA 159
QY 1601 AGCAGTTCTTCTCTGCAAAAGCAGTGGACATGAGGAAACCCAGTTCGCTATAGCAGCA 1660
Db 158 AGCAGTTCTTCTCTGCAAAAGCAGTGGACATGAGGAAACCCAGTTCGCTATAGCAGCA 99
QY 1661 GCGAAGTTAATCAGCTGAGTCCAGAGAAAGTCCACACAGTGTGCGAGGCTGACTCTGCAG 1720
Db 98 GCGAAGTTAATCAGCTGAGTCCAGAGAAAGTCCACACAGTGTGCGAGGCTGACTCTGCAG 39
QY 1721 AGTATGCTCAGCCACTGTTAGAGAAATTTGGTGTAC 1757
Db 38 AGTATGCTCAGCCACTGTTAGAGAAATTTGGTGTAC 2
RESULT 9
ABA60693
ID ABA60693 standard; DNA; 580 BP.
XX
AC ABA60693;
XX
AC
XX
DT 01-FEB-2002 (first entry)
XX
XX
DE Human foetal liver single exon nucleic acid probe #8998.
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX

PN WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00669.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 1: SEQ ID NO 8998; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;
Query Match 21.6%; Score 473.8; DB 22; Length 580;
Best Local Similarity 99.6%; Pred. No. 7.5e-138;
Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1714 TCTGCAGAGTAGTGTTCAGCCACTGGTAGGAGGAATTGTTGTACACTTCATCAAGATCT 1773
DB 64 TCCACAGAGTAGTGTTCAGCCACTGGTAGGAGGAATTGTTGTACACTTCATCAAGATCT 123
QY 1774 ACCTTTAAACCAAGAGAGAAAGACAGGCTATGACAGCTAGATGCTTACAACTCA 1833
DB 124 ACCTTTAAACCAAGAGAGAAAGACAGGCTATGACAGCTAGATGCTTACAACTCA 183
QY 1834 CCAGGCGAGGAGTTTATCATGCCCTATGCTGAACCACTCCCAATTCAGGGGCTCAGTAT 1893
DB 184 CCAGGCGAGGAGTTTATCATGCCCTATGCTGAACCACTCCCAATTCAGGGGCTCAGTAT 243
QY 1894 GCAACCCCAATCATCATGGACATGTCAGGCGACCCCACTTCAGTTGGTCAGCCCTCC 1953
DB 244 GCAACCCCAATCATCATGGACATGTCAGGCGACCCCACTTCAGTTGGTCAGCCCTCC 303
QY 1954 ACATCCACTTTCAAGGCTACGGGGAACCAACTCCCACTAGTGGGAACCTTACAACTCA 2013
DB 304 ACATCCACTTTCAAGGCTACGGGGAACCAACTCCCACTAGTGGGAACCTTACAACTCA 363
QY 2014 CTCTCTCCAGGACTGACAGTGTCTCCCTCAGCCAGGCGCCAGTATGATACCCGAAAGCT 2073
DB 364 CTCTCTCCAGGACTGACAGTGTCTCCCTCAGCCAGGCGCCAGTATGATACCCGAAAGCT 423
QY 2074 GGAAGCCAGGTTCTACCTGCCCCAGACGAATTTGGTGTACCAAGGTCACAGAGCACAA 2133
DB 424 GGAAGCCAGGTTCTACCTGCCCCAGACGAATTTGGTGTACCAAGGTCACAGAGCACAA 483
QY 2134 GAAATATCAGGAGCAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGA 2190
DB 484 GAAATATCAGGAGCAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGA 540

RESULT 10

AAK08975

ID AAK08975 standard; DNA; 580 BP.

XX

AC AAK08975;

XX

DT 05-NOV-2001 (first entry)

XX

DE Human brain expressed single exon probe SEQ ID NO: 8966.

XX

KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.

XX

OS Homo sapiens.

XX

PN WO200157275-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00667.

XX

PR 04-FEB-2000; 2000US-0180312.

XX

PR 26-MAY-2000; 2000US-0207456.

XX

PR 30-JUN-2000; 2000US-0608408.

XX

PR 03-AUG-2000; 2000US-0632366.

XX

PR 21-SEP-2000; 2000US-0234687.

XX

PR 27-SEP-2000; 2000US-0236359.

XX

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-483446/52.

XX

XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX Example 4: SEQ ID NO: 8966; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.

XX

SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;

XX

Query Match 21.6%; Score 473.8; DB 22; Length 580;

Best Local Similarity 99.6%; Pred. No. 7.5e-138;

Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1714 TCTGCAGAGTAGTGTTCAGCCACTGGTAGGAGGAATTGTTGTACACTTCATCAAGATCT 1773

DB 64 TCCACAGAGTAGTGTTCAGCCACTGGTAGGAGGAATTGTTGTACACTTCATCAAGATCT 123

QY 1774 ACCTTTAAACCAAGAGAGAAAGACAGGCTATGACAGCTAGATGCTTACAACTCA 1833

DB 124 ACCTTTAAACCAAGAGAGAGAAAGACAGGCTATGACAGCTAGATGCTTACAACTCA 183

QY 1834 CCAGGCGAGGAGTTTATCATGCCCTATGCTGAACCACTCCCAATTCAGGGGCTCAGTAT 1893

DB 184 CCAGGCGAGGAGTTTATCATGCCCTATGCTGAACCACTCCCAATTCAGGGGCTCAGTAT 243

QY 1894 GCAACCCCAATCATCATGGACATGTCAGGCGACCCCACTTCAGTTGGTCAGCCCTCC 1953

DB 244 GCAACCCCAATCATCATGGACATGTCAGGCGACCCCACTTCAGTTGGTCAGCCCTCC 303

QY 1954 ACATCCACTTTCAAGGCTACGGGGAACCAACTCCCACTAGTGGGAACCTTACAACTCA 2013

DB 304 ACATCCACTTTCAAGGCTACGGGGAACCAACTCCCACTAGTGGGAACCTTACAACTCA 363

QY 2014 CTCTCTCCAGGACTGACAGTGTCTCCCTCAGCCAGGCGCCAGTATGATACCCGAAAGCT 2073

DB 364 CTCTCTCCAGGACTGACAGTGTCTCCCTCAGCCAGGCGCCAGTATGATACCCGAAAGCT 423

QY 2074 GGAAGCCAGGTTCTACCTGCCCCAGACGAATTTGGTGTACCAAGGTCACAGAGCACAA 2133

DB 424 GGAAGCCAGGTTCTACCTGCCCCAGACGAATTTGGTGTACCAAGGTCACAGAGCACAA 483

QY 2134 GAAATATCAGGAGCAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGA 2190

DB 484 GAAATATCAGGAGCAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGA 540

```

|||||
Db 304 ACATCCACTTTCAGGCTACGGGAACCAACCTCCCCACTAGTGGGAACCTTACAATACA 363
Qy 2014 CTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGGCCAGTATGATACCCCGAAAGCT 2073
Db 364 CTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGGCCAGTATGATACCCCGAAAGCT 423
Qy 2074 GGGAGCCAGCTACCTGCCCCAGAGCAATTTGGTGTACCAAGTGGCCACAGAGCACACAA 2133
Db 424 GGGAGCCAGCTACCTGCCCCAGAGCAATTTGGTGTACCAAGTGGCCACAGAGCACACAA 483
Qy 2134 GAAGTATCAGGAGCAGGAAGGATGGGAATGTGATGTTTTAAAGAAATCCTTTGA 2190
Db 484 GAAGTATCAGGAGCAGGAAGGATGGGAATGTGATGTTTTAAAGAAATCCTTTGA 540

RESULT 11
AAK34866
ID AAK34866 standard; DNA; 580 BP.
XX AC AAK34866;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 9423.
XX KW Human: bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PF WPI; 2001-488900/53.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 9423; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX CC the probes of the invention.
XX SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;

Query Match 21.6%; Score 473.8; DB 22; Length 580;
Best Local Similarity 99.6%; Pred. No. 7.5e-138;
Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1714 TCTGCAGATATGCTCAGCCACTGCTAGGAGGAATTTGGTGTACACTTCATCAAGATCT 1773
Db 64 TCCACAGATATGCTCAGCCACTGCTAGGAGGAATTTGGTGTACACTTCATCAAGATCT 123

```

```

Qy 1774 ACCTTTAAACAGAGGAAGAAAGCAGGCTATGCAGACCTAGATCCTTACAACTCA 1833
Db 124 ACCTTTAAACAGAGGAAGAAAGCAGGCTATGCAGACCTAGATCCTTACAACTCA 183
Qy 1834 CCAGGGCAGGAAGTTTATCATGCTTATGCTGAACCACTCCCAANTTACGGGGCTGAGTAT 1893
Db 184 CCAGGGCAGGAAGTTTATCATGCTTATGCTGAACCACTCCCAANTTACGGGGCTGAGTAT 243
Qy 1894 GCACCCCAATCATCATGAGACATGTCAGGGCAGCCACCCCAACTTCAGTTGGTCAGCCCTCC 1953
Db 244 GCACCCCAATCATCATGAGACATGTCAGGGCAGCCACCCCAACTTCAGTTGGTCAGCCCTCC 303
Qy 1954 ACATCCACTTTCAAGGCTACGGGAACCAACCTCCCCACTAGTGGGAACCTTACAATACA 2013
Db 304 ACATCCACTTTCAAGGCTACGGGAACCAACCTCCCCACTAGTGGGAACCTTACAATACA 363
Qy 2014 CTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGGCCAGTATGATACCCCGAAAGCT 2073
Db 364 CTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGGCCAGTATGATACCCCGAAAGCT 423
Qy 2074 GGGAGCCAGCTACCTGCCCCAGAGCAATTTGGTGTACCAAGTGGCCACAGAGCACACAA 2133
Db 424 GGGAGCCAGCTACCTGCCCCAGAGCAATTTGGTGTACCAAGTGGCCACAGAGCACACAA 483
Qy 2134 GAAGTATCAGGAGCAGGAAGGATGGGAATGTGATGTTTTAAAGAAATCCTTTGA 2190
Db 484 GAAGTATCAGGAGCAGGAAGGATGGGAATGTGATGTTTTAAAGAAATCCTTTGA 540

RESULT 12
AAI40582
ID AAI40582 standard; DNA; 580 BP.
XX AC AAI40582;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #9268 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX PF WPI; 2001-48897/53.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 25; SEQ ID NO 9268; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful

```



```
XX PD 09-AUG-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US000657.
XX XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-483446/52.
XX XX
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX PS Example 4; SEQ ID NO: 25430; 650pp + Sequence Listing; English.
XX XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is one of the probes of the
XX CC invention.
XX XX
XX SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;
XX XX
XX XX Query Match 21.3%; Score 467; DB 22; Length 467;
XX XX Best Local Similarity 100.0%; Pred. No. 9e-136;
XX XX Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX
QY 1721 AGTATGCTCAGCCACTGGTAGGAGGAATTGTTGTACACTTCATCAAGATCTACCTTTA 1780
DB 1 AGTATGCTCAGCCACTGGTAGGAGGAATTGTTGTACACTTCATCAAGATCTACCTTTA 60
QY 1781 AACCAGAGAAGGAGAAAGACAGCTATGCAGACCTAGATCCTTACAACCTCACCAGGC 1840
DB 61 AACCAGAGAAGGAGAAAGACAGCTATGCAGACCTAGATCCTTACAACCTCACCAGGC 120
QY 1841 AGGAAGTTTATCATGCCCTATGCTGAACCACTCCCAATTACGGGCGCTCAGTATGCAACCC 1900
DB 121 AGGAAGTTTATCATGCCCTATGCTGAACCACTCCCAATTACGGGCGCTCAGTATGCAACCC 180
QY 1901 CAATCATCATGGACATGTCCAGGACACCCCACTTCCAGTGTGTCAGCCCTCCACATCCA 1960
DB 181 CAATCATCATGGACATGTCCAGGACACCCCACTTCCAGTGTGTCAGCCCTCCACATCCA 240
QY 1961 CTTTCAAGGCTACGGGGAACCAACTCCCCCACTAGTGGGAACCTTACAATACACTTCTCT 2020
DB 241 CTTTCAAGGCTACGGGGAACCAACTCCCCCACTAGTGGGAACCTTACAATACACTTCTCT 300
QY 2021 CCAGACTGACAGTGTCTTCAGCCAGGCCAGTATGATACCCGAAAGCTGGGAAGC 2080
DB 301 CCAGACTGACAGTGTCTTCAGCCAGGCCAGTATGATACCCGAAAGCTGGGAAGC 360
QY 2081 CAGGTCTACCTGCCCCAGACCAATTTGTTACCAAGGTGCCACAGACGACACAAAGTAT 2140
DB 361 CAGGTCTACCTGCCCCAGACCAATTTGTTACCAAGGTGCCACAGACGACACAAAGTAT 420
QY 2141 CAGGAGCAGGAAGGATGGGAATGTGATGTTTTTAAGAAATCCTT 2187
DB 421 CAGGAGCAGGAAGGATGGGAATGTGATGTTTTTAAGAAATCCTT 467
XX XX
XX XX RESULT 15
XX XX AAK51451
```

```
ID AAK51451 standard; DNA; 467 BP.
XX XX
AC AAK51451;
XX XX
XX DT 06-NOV-2001 (first entry)
XX XX
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 26008.
XX XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX XX
XX PD 09-AUG-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US000668.
XX XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488900/53.
XX XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX XX
XX PS Example 4; SEQ ID NO: 26008; 658pp + Sequence Listing; English.
XX XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX CC the probes of the invention.
XX XX
XX SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;
XX XX
XX XX Query Match 21.3%; Score 467; DB 22; Length 467;
XX XX Best Local Similarity 100.0%; Pred. No. 9e-136;
XX XX Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX
QY 1721 AGTATGCTCAGCCACTGGTAGGAGGAATTGTTGTACACTTCATCAAGATCTACCTTTA 1780
DB 1 AGTATGCTCAGCCACTGGTAGGAGGAATTGTTGTACACTTCATCAAGATCTACCTTTA 60
QY 1781 AACCAGAGAAGGAGAAAGACAGCTATGCAGACCTAGATCCTTACAACCTCACCAGGC 1840
DB 61 AACCAGAGAAGGAGAAAGACAGCTATGCAGACCTAGATCCTTACAACCTCACCAGGC 120
QY 1841 AGGAAGTTTATCATGCCCTATGCTGAACCACTCCCAATTACGGGCGCTCAGTATGCAACCC 1900
DB 121 AGGAAGTTTATCATGCCCTATGCTGAACCACTCCCAATTACGGGCGCTCAGTATGCAACCC 180
QY 1901 CAATCATCATGGACATGTCCAGGACACCCCACTTCCAGTGTGTCAGCCCTCCACATCCA 1960
DB 181 CAATCATCATGGACATGTCCAGGACACCCCACTTCCAGTGTGTCAGCCCTCCACATCCA 240
QY 1961 CTTTCAAGGCTACGGGGAACCAACTCCCCCACTAGTGGGAACCTTACAATACACTTCTCT 2020
DB 241 CTTTCAAGGCTACGGGGAACCAACTCCCCCACTAGTGGGAACCTTACAATACACTTCTCT 300
QY 2021 CCAGACTGACAGTGTCTTCAGCCAGGCCAGTATGATACCCGAAAGCTGGGAAGC 2080
XX XX
```

Db 301 CCAGGACTGACAGCTGCTCCTCAGCCCGAGGCCAGTATGATACCCCGAAGCTGGGAAGC 360
Qy 2081 CAGGTCTACCTGCCCCCAGACCGAATTGGTGTACCGAGTGCCCGACAGAGCCACACAAGAGTAT 2140
|||||
Db 361 CAGGTCTACCTGCCCCCAGACCGAATTGGTGTACCGAGTGCCCGACAGAGCCACACAAGAGTAT 420
Qy 2141 CAGGAGCAGGAAGGATGGGGAATGTGATGTTTAAAGAAATCCTT 2187
|||||
Db 421 CAGGAGCAGGAAGGATGGGGAATGTGATGTTTAAAGAAATCCTT 467

Search completed: January 21, 2003, 02:15:42
Job time : 456.376 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 01:41:05 ; Search time 3079.69 Seconds
(without alignments)
11516.791 Million cell updates/sec

Title: US-10-060-830-2
Perfect score: 2190
Sequence: 1 atgcctctgtctctctgcgt.....tttttaagaatacctttga 2190

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_estl:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_Other:*

26: em_gss_pro:*

27: em_gss_rod:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	837	38.2	1065	13 BM450083	BM450083 AGENCOURT
2	732	33.4	1061	14 BM905706	BM905706 AGENCOURT
C 3	684.8	31.3	712	9 AI378788	AI378788 tc67b04.x
4	683.4	21.2	694	13 BI860608	BI860608 603386804
C 5	623.2	28.5	641	9 AI433602	AI433602 th79c10.x
C 6	613.4	28.0	680	14 BQ014739	BQ014739 UI-H-ED1-

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

7	612	27.9	693	12	BF732384	BF732384 naella12.
8	590	26.9	664	10	AW383889	AW383889 QV3-HF036
9	589.8	26.9	908	12	BG178371	BG178371 602330141
C 10	584.8	26.7	671	9	AA147037	AA147037 zo32a05.s
11	584	26.7	618	12	BE894226	BE894226 601437167
12	579.6	26.5	692	12	BF732861	BF732861 naella12.
13	577.2	26.4	703	10	AW070902	AW070902 xa31B09.x
14	571.2	26.1	1272	11	AK006805	AK006805 Mus muscu
C 15	530.8	24.2	535	9	AI417256	AI417256 tg76f11.x
16	528	24.1	605	10	AW138866	AW138866 UI-H-B11-
17	518	23.7	587	10	AW383890	AW383890 QV3-HF036
18	505.8	23.1	575	10	AW383902	AW383902 QV3-HF036
19	495.4	22.6	566	12	BF434973	BF434973 7p04a07.x
C 20	495	22.6	643	10	BB179946	BB179946 BBI79946
C 21	494.8	22.6	538	14	N21309	N21309 yx53g04.s1
C 22	494	22.6	575	10	AV603144	AV603144 AV603144
C 23	490.2	22.4	560	12	BF074573	BF074573 222024 MA
24	484	22.1	600	14	BM781107	BM781107 MLN1.4.E0
25	472.8	21.6	566	9	AA147128	AA147128 zo32a05.r
26	465	21.2	526	9	AI420312	AI420312 tf06g08.x
27	464.8	21.2	912	12	BF167024	BF167024 601774167
C 28	459	21.0	459	9	AA85289	AA85289 al58g05.s
29	456.4	20.8	466	9	AI750960	AI750960 cn06g08.x
30	454.6	20.8	608	10	BB657747	BB657747 BB657747
C 31	450.6	20.6	457	9	AI276772	AI276772 ql64h03.x
32	444	20.3	459	14	N46066	N46066 yy35g05.r1
33	444	20.3	504	9	AI565996	AI565996 tn52b03.x
34	432	19.7	502	9	AI499115	AI499115 to05a05.x
C 35	430.6	19.7	439	14	H99543	H99543 yx29g09.r1
36	421.2	19.2	628	14	BQ037529	BQ037529 pgn1c.pk0
C 37	420.4	19.2	525	9	AI290500	AI290500 ql58d11.x
38	402.8	18.4	422	9	AL046164	AL046164 DK2P2434B
39	400	18.3	470	9	AI767271	AI767271 wh25f10.x
40	398.4	18.2	667	9	AL586285	AL586285 AL586285
41	398.2	18.2	455	13	BI132538	BI132538 AR032C05L
C 42	394	18.0	395	9	AI242761	AI242761 ql18h05.x
43	390.2	17.8	534	10	AV603143	AV603143 AV603143
44	384.4	17.6	421	14	N31244	N31244 yx53g04.r1
45	377.4	17.2	463	12	BF777080	BF777080 288892 MA

ALIGNMENTS

RESULT 1
BM450083
LOCUS BM450083 1065 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6393434 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528102
5', mRNA sequence.
ACCESSION BM450083
VERSION BM450083.1 GI:18499123
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1065)
NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12204 row: g column: 15
High quality sequence stop: 665.
Location/Qualifiers
1. .1065

BASE COUNT	312 a	280 c	236 g	237 t	ORIGIN
	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5528102" /clone_lib="NIH_MGC_72" /tissue_type="melanotic melanoma" /lab_host="DH10B (phage-resistant)" /notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."				

Query Match	38.2%	Score 837	DB 13	Length 1065
Best Local Similarity	97.9%	Pred. No. 8.8e-229		
Matches 859	Conservative 0	Mismatches 15	Indels 3	Gaps 1
QY 1315	CGTGCCCAAAATTTACGCAACCAGCTACAACCTGCGAGTAGCAATGAATTTCTCTGCACAG	1374		
Db	1 CGTGCCCAAAATTTACGCAACCAGCTACAACCTGCGAGTAGCAATGAATTTCTCTGCACAG	60		
QY 1375	ACAGAACAACAACCTGCCAGTCCGTATATCAGAATACTACCGTAACCTCCAATAATGTATACC	1434		
Db	61 ACAGACAAACAACCTGCCAGTCCGTATATCAGAATACTACCGTAACCTCCAATAATGTATACC	120		
QY 1435	AAAGATGAGCGCTGGCTGCGAGTTCTTGTCCTCTGTGCTGTCATGGTCTCCTCAGTACTCTC	1494		
Db	121 AAAGATGAGCGCTGGCTGCGAGTTCTTGTCCTCTGTGCTGTCATGGTCTCCTCAGTACTCTC	180		
QY 1495	ATTCTCATATTTAGTGTGTGCTTGGCACTGGAGAACAGAAAGAAAAAATCGAAGGCACC	1554		
Db	181 ATTCTCATATTTAGTGTGTGCTTGGCACTGGAGAACAGAAAGAAAAAATCGAAGGCACC	240		
QY 1555	TATGACTTTACCTTACTTGGACCGGGCAGTTGGTGGAAAGGAATGAAGCACTTTCTTCCT	1614		
Db	241 TATGACTTTACCTTACTTGGACCGGGCAGTTGGTGGAAAGGAATGAAGCACTTTCTTCCT	300		
QY 1615	GCRAAGCAGTGGACCATGAGGAAACCCAGTTCGCTATAGCAGCAGCGAAGTTAATCAC	1674		
Db	301 GCRAAGCAGTGGACCATGAGGAAACCCAGTTCGCTATAGCAGCAGCGAAGTTAATCAC	360		
QY 1675	CTGAGTCCAAGAGAGTCACACAGTCTGCAGGCTGACCTGCAGAGTATGCTCAGGCCA	1734		
Db	361 CTGAGTCCAAGAGAGTCACACAGTCTGCAGGCTGACCTGCAGAGTATGCTCAGGCCA	420		
QY 1735	CTGTAGGAGGAATTTGGTGTACACTTCATCAAAGATCTACTTTTAAACCAAGAGAAGGA	1794		
Db	421 CTGTAGGAGGAATTTGGTGTACACTTCATCAAAGATCTACTTTTAAACCAAGAGAAGGA	480		
QY 1795	AAAGAAGCAGCTATGCAGACCTAGATCTTTACAACTCACCGGCGCAGGAAGTTATCAT	1854		
Db	481 AAAGAAGCAGCTATGCAGACCTAGATCTTTACAACTCACCGGCGCAGGAAGTTATCAT	540		
QY 1855	GCCTATGCTGAACACATCCCAATTTACGGGGCTGAGTATGCAACCCCAATCATATGGAC	1914		
Db	541 GCCTATGCTGAACACATCCCAATTTACGGGGCTGAGTATGCAACCCCAATCATATGGAC	600		
QY 1915	ATGTGAGGGACCCCAACAATTTCAAGTTGGTCAGCCCTCCACATCCACTTTCAAGGCTACG	1974		
Db	601 ATGTGAGGGACCCCAACAATTTCAAGTTGGTCAGCCCTCCACATCCACTTTCAAGGCTACG	660		
QY 1975	GGGAACCAACCTCCCCCAGTGTGGGAACCTTACAATACACTTCTCTCGAGGACTGCACAGC	2034		
Db	661 GGGAACCAACCTCCCCCAGTGTGGGAACCTTACAATACACTTCTCTCGAGGACTGCACAGC	720		
QY 2035	TGCTCTTCAGCCAGGCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGTCTACCTGCC	2094		
Db	721 TGCTCTTCAGCCAGGCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGTCTACCTGCC	780		
QY 2095	CCAGACCAATTTGGTGTACCAGGTGCCACAGCAGCACAAAGATATCAGGAGCAGGAAG	2154		
Db	781 CCAGACCAATTTGGTGTACCAGGTGCCACAGCAGCACAAAGATATCAGGAGCAGGAAG	840		

Qy	2155	G---ATGGGGAATGTGATGTTTAAAGAAATCCTTT	2188
Db	841	GGATCGGGAATGTGAATGGTTTAAAGAAATCCTT	877
RESULT 2			
BM905706			
LOCUS			
DEFINITION	BM905706	1061 bp	linear
ACCESSION	AGENCOURT_6721242	NIH_MGC_71	Homo sapiens
VERSION	5,	cdna clone	IMAGE:5556057
KEYWORDS	BM905706	sequence.	
SOURCE	BM905706.1	GI:19356085	
ORGANISM	EST.	human.	
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	1 (bases 1 to 1061)		
COMMENT	NIH-MGC http://mgc.nci.nih.gov/ .		
	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabs-rc@mail.nih.gov		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LLAM12277	row: d	column: 10
	High quality sequence stop: 629.		
FEATURES	Location/Qualifiers		
source	1..1061		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5556057"		
	/clone_lib="NIH_MGC_71"		
	/tissue_type="leiomyosarcoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: uterus; Vector: PCMV-SF0RT6; Site_1: NotI;		
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.		
	Average insert size 2.1 kb.		
BASE COUNT	281 a	255 g	283 t
ORIGIN	241 c	1	others
Query Match	33.4%	Score 732;	DB 14; Length 1061;
Best Local Similarity	94.0%;	Prod. No. 1.3e-198;	
Matches 829; Conservative	0;	Mismatches 40;	Indels 13; Gaps 6;

[illegible]

[illegible]

```
||||| 580 TAACAGGCATTATACCCACTGGAT--ACCATGTTGGAGCACAATTACTATGTGTCTGCCT 523
QY 1004 ACAGATCCCTGTACAGTATGATGGGAGAAATGGAGCTGTGTACAGAGAGCCGTGTGG 1063
Db 522 ACAGATCCCTGTACAGTATGATGGGAGAAATGGAGCTGTGTACAGAGAGCCGTGTGG 463
QY 1064 AGCAAGATAAGATATTTCAAGAGAAACAAAGATTATCACAGGAGTGTGCGTAATAACTTTT 1123
Db 462 AGCAAGATAAGATATTTCAAGAGAAACAAAGATTATCACAGGAGTGTGCGTAATAACTTTT 403
QY 1124 TGCACCAATATGACGCTGCTGGATGTTATTAGATGAATCTACCAATGCCAGAGAAATTTG 1183
Db 402 TGCACCAATATGACGCTGCTGGATGTTATTAGATGAATCTACCAATGCCAGAGAAATTTG 343
QY 1184 CCATGAAATGGAGCTGCTGGATGTCAGTATTCTTAAAGTCGTCCTCCAAACTTA 1243
Db 342 CCATGAAATGGAGCTGCTGGATGTCAGTATTCTTAAAGTCGTCCTCCAAACTTA 283
QY 1244 CTCACCTCCACCTCCCTCGGAACAGCAATGACCTCAAAAACACATACAGCCCTCCAAAA 1303
Db 282 CTCACCTCCACCTCCCTCGGAACAGCAATGACCTCAAAAACACATACAGCCCTCCAAAA 223
QY 1304 TAGCCAAAGTGTGCCCCAAAATTTACGCAACACATACACCTCGCAGTAGCAATGAAT 1363
Db 222 TAGCCAAAGTGTGCCCCAAAATTTACGCAACACATACACCTCGCAGTAGCAATGAAT 163
QY 1364 TTCCTGCACAGACAGACAAACAACTGCCAGTCTGTATACAGAAATACCTACCTTAAC 1423
Db 162 TTCCTGCACAGACAGACAAACAACTGCCAGTCTGTATACAGAAATACCTACCTTAAC 103
QY 1424 CAATGTAAACCAAGATGTAGCGTGGCTGAGTCTTGTGCTGCTGCTGCTGCTGCTGCTG 1483
Db 102 CAATGTAAACCAAGATGTAGCGTGGCTGAGTCTTGTGCTGCTGCTGCTGCTGCTGCTG 43
QY 1484 TCACCTACTCTCATCTCATATTAAGTGTGCTTGGCACTGGA 1525
Db 42 TCACCTACTCTCATCTCATATTAAGTGTGCTTGGCACTGGA 1

RESULT 6
BQ014739/c 680 bp mRNA linear EST 26-MAR-2002
LOCUS
DEFINITION UI-H-ED1-act-n-12-0-UI.s1 NCI_CGAP_ED1 Homo sapiens cDNA clone
IMAGE:5833547 3', mRNA sequence.
ACCESSION BQ014739
VERSION BQ014739
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rcap@rmail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLYA=yes.
Location/Qualifiers
1. 680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5833547"
/clone_lib="NCI_CGAP_ED1"
```

```
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/notes="Organ: Left Pubic Bone; Vector: pT73-Pac
(Pharmacia) with a modified polylinker; Site:1: EcoR I;
Site_2: Not I; NCI_CGAP_ED1 is a normalized cDNA library
containing the following tissue(s): Chondrosarcoma cell
line C55. The library was constructed according to Bonaldo
, Lennon and Soares, Genome Research, 6:791-806, 1996.
First strand cDNA synthesis was primed with an oligo-dT
primer containing a Not I site. Double stranded cDNA was
ligated to an EcoR I adaptor, digested with Not I, and
cloned directionally into pT73-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is GCTCAAGGCT.
TAG_LIB=UI-H-ED1
TAG_TISSUE=chondrosarcoma
TAG_SEQ=CGTCAAGGCT"
BASE COUNT 138 a 157 c 172 g 213 t
ORIGIN
Query Match 28.08; Score 613.4; DB 14; Length 680;
Best Local Similarity 99.88; Pred. No. 1.3e-164;
Matches 614; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1576 CGGCGAGTGTGGTGAAGGAATGAAGCAGTTCTCTCTCAAAAGCAGTGGACCATGAG 1635
Db 680 CGGCGAGTGTGGTGAAGGAATGAAGCAGTTCTCTCTCAAAAGCAGTGGACCATGAG 621
QY 1636 GAAACCCAGTTCCTATAGCAGCAGCAAGTAAATCACTGAGTCCAAGAGAGTCAACC 1695
Db 620 GAAACCCAGTTCCTATAGCAGCAGCAAGTAAATCACTGAGTCCAAGAGAGTCAACC 561
QY 1696 ACAGTGTGAGGCTGACTCTGAGAGATATGCTCAGCAGTGTGAGGAGGATTTGTTGGT 1755
Db 560 ACAGTGTGAGGCTGACTCTGAGAGATATGCTCAGCAGTGTGAGGAGGATTTGTTGGT 501
QY 1756 ACATTTCATCAAGATCTACCTTTAAACCAAGAGAAAGAGAGAGGCTATGCGAGAC 1815
Db 500 ACATTTCATCAAGATCTACCTTTAAACCAAGAGAAAGAGAGAGGCTATGCGAGAC 441
QY 1816 CTAGATCCTTAACTACCTACAGGAGCAAGTATATCATGCTATGCTGAACCACTCCCA 1875
Db 440 CTAGATCCTTAACTACCTACAGGAGCAAGTATATCATGCTATGCTGAACCACTCCCA 381
QY 1876 ATTACGGGCGCTGAGTATGCAACCCCAATCATATGACATGTGAGGCGACCCCACT 1935
Db 380 ATTACGGGCGCTGAGTATGCAACCCCAATCATATGACATGTGAGGCGACCCCACT 321
QY 1936 TCAGTTGCTAGCCCTCCACATCCACTTTCAAGGCTAGGGGAAACCAACCTCCCACTA 1995
Db 320 TCAGTTGCTAGCCCTCCACATCCACTTTCAAGGCTAGGGGAAACCAACCTCCCACTA 261
QY 1996 GTGGGAAGTACATACACTTCTCTCAGGAGTACAGCTGCTCTCAGCCCGAGCCAG 2055
Db 260 GTGGGAAGTACATACACTTCTCTCAGGAGTACAGCTGCTCTCAGCCCGAGCCAG 201
QY 2056 TATGATACCCGAGCTGGGAGCCAGTCTACTCTGCCAGAGCAATTTGTTGATACAG 2115
Db 200 TATGATACCCGAGCTGGGAGCCAGTCTACTCTGCCAGAGCAATTTGTTGATACAG 141
QY 2116 GTGCCACAGACACAAAGAGTATCAGGAGCAGGAAGGATGGGAAATGTGATGTTTTT 2175
Db 140 GTGCCACAGACACAAAGAGTATCAGGAGCAGGAAGGATGGGAAATGTGATGTTTTT 81
QY 2176 AAAGAAATCCTTTGA 2190
Db 80 AAAGAAATCCTTTGA 66
RESULT 7
```


LOCUS AA147037 .671 bp mRNA linear EST 05-DEC-1999
DEFINITION zoz32a05.sl Stragatene colon (#937204) Homo sapiens cDNA clone IMAGE:588560 -3' similar to TR:G704441 G704441 HYPOTHETICAL 40.0 KD PROTEIN ; , mRNA sequence.
ACCESSION AA147037
VERSION AA147037.1 GI:1716444
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 671)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissole,S., Dietrich,N., DuBouque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Nardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasaki,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra.M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT Contact: Wilison RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amer sham
High quality sequence stop: 439.

FEATURES

source	Location/Qualifiers
	1..671
	/organism="Homo sapiens"
	/db_xref="GBD:4620862"
	/db_xref="taxon:9606"
	/clone="IMAGE:588560"
	/clone_lib="Stratagene colon (#937204)"
	/tissue_type="tumor"
	/cell_line="T84 carcinoma cell line"
	/note=Organ:"SOLR cells (kanamycin resistant)"
	EcoRI; Site:-2 XhoI; Cloned unidirectionally. Primer: Oligo dt. T-84 clonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCAGCAG 3'-3' adaptor sequence: 5' CTCGGAGTTTTTTTCTTTTCTTTT 3'

```

BASE COUNT      176 a   162 c    133 g     198 t       2 others
ORIGIN
Query Match          26.7%; Score 584.8; DB 9; Length 671;
Best Local Similarity 97.9%; Pred No. 2.le-156;
Matches 645; Conservative 0; Mismatches 8; Indels 6; Gaps 5;
```

QY	515	CITTTTGCTGAGATAATCGGAACAATTCCATCGGATATAGAGATTCTCGCATGTGTCA	574
Dd	.653	CTTTTCTGAGATATCTGG-Accatttcctcgtgat-gacattcctgcgcat--gtcn	598
QY	575	TGCGTGGTGTGCATGCAGAGTAGTGCAAACACGTTGGCGGCCAAATCACTGTTGTAA	634
Dd	597	TGCGTGGTGT-CATGCAGAGTAGTGTAACACACGTT-GGCGGCCAAATCACTGTTGTAA	540
QY	635	TTAGTAAAGGTATTCCTATTATGAAGTCTTTGGCTAACACGTCACATCTGGTGTTG	694
Dd	539	TTAGTAAAGGTATTCCTATTATGAAGTCTTTGGCTAACACGTCACATCTGGTGTTG	480
QY	695	GACTTATCTACAAGTCTTTTACATTFAAGACAAGTGGATGTTATGAACACTGGGA	754
Dd	479	GACACTTATCTACAAGTCTTTTACATTFAAGACAAGTGGATGTTATGAACACTGGGA	420
QY	755	TGGAGTCTGCTGTATCGGGATCTCCAATPAACAGCATCATCTGCTGGAGTGACTG	814
Dd	419	TGGAGTCTGCTGTATCGGGATCTCCAATPAACAGCATCATCTGCTGGAGTGACTG	360

QY	361	TTCTGTTTCATGAGTGGGAATCAATGTTCTGGACGCGGATTTTGGCCCTCATCTCTGTT	420
Db	431	TTCTGTTTCATGAGTGGGAATCAATGTTCTGGACGCGGATTTTGGCCCTCATCTCTGTT	490
QY	421	ATAGATAAACAAGATCTAAATTACTTGTGTTGGACACATGCATCCCAATTTTGGAAACCTGAG	480
Db	491	ATAGATAAACAAGATCTAAATTACTTGTGTTGGACACATGCATCCCAATTTTGGAAACCTGAG	550
QY	481	TTTCAGTAAAGTACTGCCACAGCTGGTGTCTGCTCTCCCTTTTGTCTGAGATATCTGGAAACAAT	540
Db	551	TTTCAGTAAAGTACTGCCACAGCTGGTGTCTGCTCTCCCTTTTGTCTGAGATATCTGGAAACAAT	610
QY	541	CCTCATGATATAGAGATTCCTCGCCATGTGTGATGCTGGTGGTG-CATGACGAGTAGT	599
Db	611	CCTCATGATATAGAGATTCCTCGCCATGTGTGATGCTGGTGGTG-CATGACGAGTAGT	670
QY	600	GTCAAAACAGTGTGGGGCGGCAAAATCA	625
Db	671	GTCAAAACAGTGTGGGGCGGCAAAATCA	696
RESULT	14		
AC006805			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			</

KEYWORDS
SOURCE

EST. human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 535)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 452.

FEATURES
source

1. .535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2114733"
/clone_lib="Soares_NhMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBHM, pregnant uterus NBHPU, and fetal heart NBH19w) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."
BASE COUNT 133 a 105 c 132 g 164 t 1 others
ORIGIN

Query Match 24.2%; Score 530.8; DB 9; Length 535;
Best Local Similarity 99.4%; Pred. No. 6.le-141;
Matches 532; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 991 TATGTGTCCTACAGATCCTCTACAGTATGATGGGAGAAATGGACTGTGTACAGA 1050
DB 535 TATGTGTCGTGTCAGATTCCTCTACAGTATGATGGGAGAAATGGACTGTGTACAGA 476
QY 1051 GAGCTGTGTGGAGAGATAGATATTCAGGAACAAGATATTCACGAGGATGTG 1110
DB 475 GAGCTGTGTGGAGAGATAGATATTCAGGAACAAGATATTCACGAGGATGTG 416
QY 1111 CGTAATACTTTTGGCCACCAATTATTGCACGTTTTATTAGAGTGAATCCTACCCAATGG 1170
DB 415 CGTAATACTTTTGGCCACCAATTATTGCACGTTTTATTAGAGTGAATCCTACCCAATGG 356
QY 1171 CAGAGAAATGCCATGAAATGAGCTGCTCGATGTCTGTTTATTCCTAAAGTCGT 1230
DB 355 CAGCAGAAATGGCATGAAATGAGCTGCTCGATGTCTGTTTATTCCTAAAGTCGT 296
QY 1231 CCTCAAACTTACTCAACCTCCACCTCTCGGAACAGCATGACCTCAAAACACTACA 1290
DB 295 CCTCAAACTTACTCAACCTCCACCTCTCGGAACAGCATGACCTCAAAACACTACA 236
QY 1291 GCCCTCCAAAATAGCCAAAGGTGTCGCCCAAAATTTAGCAACCACTACAACCTCGC 1350
DB 235 GCCCTCCAAAATAGCCAAAGGTGTCGCCCAAAATTTAGCAACCACTACAACCTCGC 176
QY 1351 AGTAGCAATGATTTCTTCGACAGACAGACAAACAACTGCCAGTCTGTATCAGAAAT 1410
DB 175 AGTAGCAATGATTTCTTCGACAGACAGACAAACAACTGCCAGTCTGTATCAGAAAT 116
QY 1411 ACTACCGTAACCTCCAAATGTACCAAGATGTAGCGTGGCTGCAGTCTTGTCCCTGTG 1470

Db 115 ACTACCGTAACTCCAAATGTAAACAAAGATGTAGCGCTGGCTGCAGTTCTTGTCCCTGTG 56
QY 1471 CTGGTCATGGTCTCTCACTACTCTCATTTCTCATATTAGTGTGTGCTGGCACTGGA 1525
Db 55 CTGGTCATGGTCTCTCACTACTCTCAATCTCATATTAGTGTGTGCTGGCACTGGA 1
Search completed: January 21, 2003, 09:21:18
Job time : 3087.69 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 00:35:34 ; Search time 5604.73 Seconds
(without alignments)
11371.681 Million cell updates/sec

Title: US-10-060-830-2

Perfect score: 2190

Sequence: 1 atgcctctgtctctctgct.....ttttaagaagaatcctttga 2190

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vt.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vt.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2190	100.0	2939	9	AF387547 Homo sapi
2	2188.4	99.9	6093	9	AB073146 Homo sapi
3	2133.2	97.4	3858	9	BC029658 Homo sapi
4	1669.8	76.2	2339	10	AF387549 Rattus no
5	1663.6	76.0	2642	10	AF387548 Mus muscu
6	1521.2	69.5	2020	9	AK056350 Homo sapi
7	1451.8	66.3	1907	6	AX405795 Sequence
8	1158.4	52.9	1388	9	HUMCUB1 D29810 Human mRNA
9	473.8	21.6	192247	2	AC013497 Homo sapi
10	473.8	21.6	193623	9	AC106728 Homo sapi
11	462.8	21.1	146515	2	AC025661 Homo sapi
12	370.4	16.9	586	9	BC007117 Homo sapi
13	353.8	16.2	203102	2	AC126975 Rattus no
14	232.4	10.6	164216	2	AC091213 Homo sapi
15	224.2	10.2	1620	6	AX118820 Sequence
16	224.2	10.2	1761	6	AX118818 Sequence
17	224.2	10.2	1768	6	AX118822 Sequence
18	222.6	10.2	2547	9	AK095973 Homo sapi
19	218	10.0	192247	2	AC013497 Homo sapi
20	204	9.3	1464	6	AX118816 Sequence
21	198.8	9.1	59871	2	AC101046 Mus muscu
22	174.8	8.0	217	6	AX182030 Sequence
23	157.4	7.2	61344	2	AC073443 Homo sapi
24	142.4	6.5	164216	2	AC091213 Homo sapi
25	135.4	6.2	146515	2	AC025661 Homo sapi
26	135.2	6.2	59871	2	AC101046 Mus muscu
27	131.2	6.0	61344	2	AC073443 Homo sapi
28	130.4	6.0	203102	2	AC126975 Rattus no
29	119	5.4	2846	10	BC026771 Mus muscu
30	94.4	4.3	6895	4	BOVFACV2A M81441 Bos taurus
31	94.4	4.3	6910	4	BOVFACV S80643 Bos taurus
32	92.8	4.2	1895	4	BTPAS67PT X91895 B.taurus MR
33	92.8	4.2	1965	4	BTPAS67PT U52925 Mus musculus
34	91.6	4.2	6585	10	MMU52925 AF191308 Sus scrofa
35	91.6	4.2	7062	4	HSU70312 U70312 Homo sapien
36	89.8	4.1	1719	9	BC030828 Homo sapi
37	89.8	4.1	2728	9	BC030828 Homo sapi
38	89.8	4.1	4681	6	AX473349 Sequence
39	88.8	4.1	1227	4	ETBP47PRO Y11719 Bos taurus
40	88.2	4.0	6893	9	HUMEVA M14335 Human coagu
41	88	4.0	2303	10	AF031524 Mus muscu
42	86.6	4.0	6909	6	A46255 Sequence 1
43	86.6	4.0	6909	6	A63218 Sequence 1
44	86.6	4.0	6909	6	AR071379 Sequence
45	86.6	4.0	6909	6	AX146885 Sequence

ALIGNMENTS

RESULT 1

AF387547

LOCUS

DEFINITION

2939 bp mRNA linear PRI 11-NOV-2001

Homo sapiens endothelial and smooth muscle cell-derived

neuropilin-like protein (ESDN) mRNA, complete cds.

AF387547

ACCESSION

AF387547

VERSION

AF387547.1

KEYWORDS

GI:16902434

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2939)

Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,

Matsumori, A., Sasayama, S., Honjo, T. and Tashiro, K.

QY 181 ATGGGAGAGAGATTTCGCATCAAAATTTGGTGACTTTGACATTGAGATTCTGATTCCTGT 240
|||||
Db 682 ATGGGAGAGAGATTTCGCATCAAAATTTGGTGACTTTGACATTGAGATTCTGATTCCTGT 741
|||||
QY 241 CACTTTAATTACTTTGAGAAATTTATATGAATTTGGAGTCAGCAGAACTGAAATAGGCAAA 300
|||||
Db 742 CACTTTAATTACTTTGAGAAATTTATATGAATTTGGAGTCAGCAGAACTGAAATAGGCAAA 801
|||||
QY 301 TACTGTGCTGGTGGTTCGAATGAACCAATCAATTTGAATCAAAAGGCAATGAATACACA 360
|||||
Db 802 TACTGTGCTGGTGGTTCGAATGAACCAATCAATTTGAATCAAAAGGCAATGAATACACA 861
|||||
QY 361 TTGCTGTTTCATGAGTGGAAATCATCTTTCTGGACGCGATTTTGGCCCTCATACTCTCTT 420
|||||
Db 862 TTGCTGTTTCATGAGTGGAAATCATCTTTCTGGACGCGATTTTGGCCCTCATACTCTCTT 921
|||||
QY 421 ATAGATAAACAAGATCTAATTACTTTGTTGGACACTGTCATCAATTTTGGAACTCTGAG 480
|||||
Db 922 ATAGATAAACAAGATCTAATTACTTTGTTGGACACTGTCATCAATTTTGGAACTCTGAG 981
|||||
QY 481 TTCAGTAACTAGTCCAGCTGGTGTCTGCTTTCTTTGCTGAGATATCTGGAACTAT 540
|||||
Db 982 TTCAGTAACTAGTCCAGCTGGTGTCTGCTTTCTTTGCTGAGATATCTGGAACTAT 1041
|||||
QY 541 CCTCATGATATAGAGATTCCTCGCATTTGTCATGCTGTGTGTCATGCGAGGAGTAGTG 600
|||||
Db 1042 CCTCATGATATAGAGATTCCTCGCATTTGTCATGCTGTGTGTCATGCGAGGAGTAGTG 1101
|||||
QY 601 TCACAACGTTGGGGGCGCAATCACTAGTGTGTAAATAGTAAAGTATTCCTATTATGAA 660
|||||
Db 1102 TCACAACGTTGGGGGCGCAATCACTAGTGTGTAAATAGTAAAGTATTCCTATTATGAA 1161
|||||
QY 661 AGTTCTTTGGCTAAACAAGTACATCTGTGGTGGACACTTATCTACAAAGTCTTTTACA 720
|||||
Db 1162 AGTTCTTTGGCTAAACAAGTACATCTGTGGTGGACACTTATCTACAAAGTCTTTTACA 1221
|||||
QY 721 TTTAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATCGCGATCCT 780
|||||
Db 1222 TTTAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATCGCGATCCT 1281
|||||
QY 781 CAAATAACAGCATATCTGTGCTGGAGTGGACTGACACACAGGCGAAGAAACAGTTGG 840
|||||
Db 1282 CAAATAACAGCATATCTGTGCTGGAGTGGACTGACACACAGGCGAAGAAACAGTTGG 1341
|||||
QY 841 AAACCCAAAAGCCAGCTGAAACAACTGGACCGCTTGGCTGCTTTGCCACTGAT 900
|||||
Db 1342 AAACCCAAAAGCCAGCTGAAACAACTGGACCGCTTGGCTGCTTTGCCACTGAT 1401
|||||
QY 901 GAATACCAAGTGTACAAATAGATTGAATAGGAAAAAGAAATAACAGGCATTATAACC 960
|||||
Db 1402 GAATACCAAGTGTACAAATAGATTGAATAGGAAAAAGAAATAACAGGCATTATAACC 1461
|||||
QY 961 ACTGATCCACCATGGTGGAGCACAATTTACTATGTCTGCTGCTACAGAACTCTGTACAGT 1020
|||||
Db 1462 ACTGATCCACCATGGTGGAGCACAATTTACTATGTCTGCTGCTACAGAACTCTGTACAGT 1521
|||||
QY 1021 GATGATGGCAGAAATGCACTGTGTACAGAGAGCTGTGTGGAGCAAGATAAGATATT 1080
|||||
Db 1522 GATGATGGCAGAAATGCACTGTGTACAGAGAGCTGTGTGGAGCAAGATAAGATATT 1581
|||||
QY 1081 CAAGGAAACAAAGATTATACACAGGATGGTAAATACTTTTGGCCACCAATTTATGCA 1140
|||||
Db 1582 CAAGGAAACAAAGATTATACACAGGATGGTAAATACTTTTGGCCACCAATTTATGCA 1641
|||||
QY 1141 CGTTTATTAGAGTGAATCCTACCANTGGCAGCAGAAATTTGCCATGAAATGGAGCTG 1200
|||||
Db 1642 CGTTTATTAGAGTGAATCCTACCANTGGCAGCAGAAATTTGCCATGAAATGGAGCTG 1701
|||||
QY 1201 CTCGGATGTCAGTTTATCTTAAAGTGTCTCTCCAAACTTACTCAACCTCCACCTCCT 1260
|||||
Db 1702 CTCGGATGTCAGTTTATCTTAAAGTGTCTCTCCAAACTTACTCAACCTCCACCTCCT 1761
|||||
QY 1261 CGGAAACAGCAATGACCTCAAAAACACTACAGCCCTTCCAAAAATAGCCAAAGGTGCTGCC 1320
|||||

Db 1762 CGGAACAGCAATGACCTCAAAAACACTACAGCCCTTCCAAAATAGCCAAAGGTGCTGCC 1821
|||||
QY 1321 CCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATGAATTTCTTCGACAGACAGAA 1380
|||||
Db 1822 CCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATGAATTTCTTCGACAGACAGAA 1881
|||||
QY 1381 CAACAACAGTGCAGCTCTGATATCAGAAATACTACCGTAACCTCCAAATGTACCAAGAT 1440
|||||
Db 1882 CAACAACAGTGCAGCTCTGATATCAGAAATACTACCGTAACCTCCAAATGTACCAAGAT 1941
|||||
QY 1441 GTAGCGTGGCTGCAAGTTCTTGTCTGCTGTGGTGTGCTGCTCCTCCTACTCTCATCTCTC 1500
|||||
Db 1942 GTAGCGTGGCTGCAAGTTCTTGTCTGCTGTGGTGTGCTGCTCCTCCTACTCTCATCTCTC 2001
|||||
QY 1501 ATATTAGTGTGTGCTTGGCACTGGAGAACAGAAAGAAAAAACTGAAGGCACCTATGAC 1560
|||||
Db 2002 ATATTAGTGTGTGCTTGGCACTGGAGAACAGAAAGAAAAAACTGAAGGCACCTATGAC 2061
|||||
QY 1561 TTACTTACTGGGACCGGCGAGGTTGGTGGAAAGGAATGAAGCAGTCTTCTCTCGCAAA 1620
|||||
Db 2062 TTACTTACTGGGACCGGCGAGGTTGGTGGAAAGGAATGAAGCAGTCTTCTCTCGCAAA 2121
|||||
QY 1621 GCAGTGGACCATGAGGAAACCCCACTTCGTATAGCAGCAGGAAAGTTAATCACCTGAGT 1680
|||||
Db 2122 GCAGTGGACCATGAGGAAACCCCACTTCGTATAGCAGCAGGAAAGTTAATCACCTGAGT 2181
|||||
QY 1681 CCAAGAGAAGTCAACACAGTGTGCAGGCTGACTCTGCAGAGATATGCTCAGCCACTGGTA 1740
|||||
Db 2182 CCAAGAGAAGTCAACACAGTGTGCAGGCTGACTCTGCAGAGATATGCTCAGCCACTGGTA 2241
|||||
QY 1741 GGAGGAATTTGGTACACTTCATCAAGATCTACTTTAAACCAGAGAAAGAAAGAA 1800
|||||
Db 2242 GGAGGAATTTGGTACACTTCATCAAGATCTACTTTAAACCAGAGAAAGAAAGAA 2301
|||||
QY 1801 GCAGCTATGCAGACCTAGATCTTACAACTCACACAGGCGAGAAAGTTTATCATGCTAT 1860
|||||
Db 2302 GCAGCTATGCAGACCTAGATCTTACAACTCACACAGGCGAGAAAGTTTATCATGCTAT 2361
|||||
QY 1861 GCTGAACACCTCCCAATTTACGGGGCTGAGTATGCAACCCCAATCATCATGACATGCA 1920
|||||
Db 2362 GCTGAACACCTCCCAATTTACGGGGCTGAGTATGCAACCCCAATCATCATGACATGCA 2421
|||||
QY 1921 GGGCACCCCACTTCAGTTGCTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAAC 1980
|||||
Db 2422 GGGCACCCCACTTCAGTTGCTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAAC 2481
|||||
QY 1981 CAACCTCCCCCACTAGTGGGAACCTTACAATACACTTCTCTCCAGACTGACAGCTGCTCC 2040
|||||
Db 2482 CAACCTCCCCCACTAGTGGGAACCTTACAATACACTTCTCTCCAGACTGACAGCTGCTCC 2541
|||||
QY 2041 TCAGCCAGGCGCCAGTATGATACCCGGAAGCTGGGAAGCCAGGCTTACCTGCCCCAGAC 2100
|||||
Db 2542 TCAGCCAGGCGCCAGTATGATACCCGGAAGCTGGGAAGCCAGGCTTACCTGCCCCAGAC 2601
|||||
QY 2101 GAAATGGTGTACCAGTCCACAGAGCACACAAGAAATATCAGGAGCAGAAAGGATGG 2160
|||||
Db 2602 GAAATGGTGTACCAGTCCACAGAGCACACAAGAAATATCAGGAGCAGAAAGGATGG 2661
|||||
QY 2161 GAAATGGTGTATTTTAAAGAAATCCTTTGA 2190
|||||
Db 2662 GAAATGGTGTATTTTAAAGAAATCCTTTGA 2691
|||||

RESULT 3

BC029658

LOCUS

DEFINITION

BC029658 3858 bp mRNA linear PRI 20-MAY-2002
Homo sapiens, similar to endothelial and smooth muscle cell-derived
neuropilin-like protein, clone MGC:30023 IMAGE:4431275, mRNA,
complete cds.

ACCESSION

BC029658

VERSION

BC029658.1

KEYWORDS

MGC.

QY	361	TTGCTGTTTCATGAGTGGAAATCCATGTTTCTTGGACGGGGATTTTGGCCCTCATACTCTGTT	420
Db	855	TTGCTGTTTCATGAGTGGAAATCCATGTTTCTTGGACGGGGATTTTGGCCCTCATACTCTGTT	914
QY	421	ATAGATAAACACAGATCTAAATTACTTGTGTCGACACTGCATCCAAATTTTGGAACTGAG	480
Db	915	ATAGATAAACACAGATCTAAATTACTTGTGTCGACACTGCATCCAAATTTTGGAACTGAG	974
QY	481	TTCACTAAGTACTGCCACGTGGTGTCTGCTCTTCTTGTCTGAGATATCTCGACAAT	540
Db	975	TTCACTAAGTACTGCCACGTGGTGTCTGCTCTTGTCTGAGATATCTCGACAAT	1034
QY	541	CCTCATGGATATAGAGATTCTCGCCATTGTGATGGCTGGTGTGATCGACGAGTAGTG	600
Db	1035	CCTCATGGATATAGAGATTCTCGCCATTGTGATGGCTGGTGTGATCGACGAGTAGTG	1094
QY	601	TCAAACACGTTTGGCGGCCAAATCAGTTGTAAATTAGTAAAGGTATCCCTTATTATGAA	660
Db	1095	TCAAACACGTTTGGCGGCCAAATCAGTTGTAAATTAGTAAAGGTATCCCTTATTATGAA	1154
QY	661	AGTTCTTTGGCTAAACAAGTCAACATCTGTGGTGGGACCTTATCTACAGTCTTTTTACA	720
Db	1155	AGTTCTTTGGCTAAACAAGTCAACATCTGTGGTGGGACCTTATCTACAGTCTTTTTACA	1214
QY	721	TTTAAAGCAAGTGGATGTTTATGGAACACTGGGGATGGAGCTGTGGTGTATCGCGATCCT	780
Db	1215	TTTAAAGCAAGTGGATGTTTATGGAACACTGGGGATGGAGCTGTGGTGTATCGCGATCCT	1274
QY	781	CAATATACAGCATCATCTGTCTGGAGTGGACTGACCCACAGGGCAAGAACAGTTGG	840
Db	1275	CAATATACAGCATCATCTGTCTGGAGTGGACTGACCCACAGGGCAAGAACAGTTGG	1334
QY	841	AAACCCAAAAAGCCAGGCTGAAAAACCTGGACCGCCTTGGGCTCTTTTGGCCACTGAT	900
Db	1335	AAACCCAAAAAGCCAGGCTGAAAAACCTGGACCGCCTTGGGCTCTTTTGGCCACTGAT	1394
QY	901	GAATACCAAGTGGTTTACAAATAGATTTGAATAAGGAAAGAAAATAACAGGCAATTATACC	960
Db	1395	GAATACCAAGTGGTTTACAAATAGATTTGAATAAGGAAAGAAAATAACAGGCAATTATACC	1454
QY	961	ACTGGATCCACCATGTTGGAGCACAATTACTATGTGTCTGCTACAGAAATCTGTACAGT	1020
Db	1455	ACTGGATCCACCATGTTGGAGCACAATTACTATGTGTCTGCTACAGAAATCTGTGACGT	1514
QY	1021	GATGATGGCAGAAATGGACTGTGTACAGAGAGCCTGGTGGAGCAAGATAAGATATTT	1080
Db	1515	GATGATGGCAGAAATGGACTGTGTACAGAGAGCCTGGTGGAGCAAGATAAGATATTT	1574
QY	1081	CAAGAAACAAAGATTATCACCAGGATGTCGCTAATAACTTTTTGCCACCAATTTATGCA	1140
Db	1575	CAAGAAACAAAGATTATCACCAGGATGTCGCTAATAACTTTTTGCCACCAATTTATGCA	1634
QY	1141	CGTTTTATTAGCTGAATCCTACCCCAATGGCAGCAGAAAATGGCATGAAAATGGAGCTG	1200
Db	1635	CGTTTTATTAGCTGAATCCTACCCCAATGGCAGCAGAAAATGGCATGAAAATGGAGCTG	1694
QY	1201	CTCGGATGTCAGTTTATCTTAAAGTCTGCTCTCCAAAACCTTACTCAACCTCCACCTCCT	1260
Db	1695	CTCGGATGTCAGTTTATCTTAAAGTCTGCTCTCCAAAACCTTACTCAACCTCCACCTCCT	1754
QY	1261	CGGACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAATAGCCAAAGTCGTGCC	1320
Db	1755	CGGACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAATAGCCAAAGTCGTGCC	1814
QY	1321	CCAAAATTTACGCAACCACTACAACCTCGCAGTAGCATGAATTTCTGTGCACAGACAGAA	1380
Db	1815	CCAAAATTTACGCAACCACTACAACCTCGCAGTAGCATGAATTTCTGTGCACAGACAGAA	1874
QY	1381	CAAAACACTGCCAGTCTCTGATATCAGAAATACTACCGTAACTCCAAATGTAAACCAAGAT	1440
Db	1875	CAAAACACTGCCAGTCTCTGATATCAGAAATACTACCGTAACTCCAAATGTAAACCAAGAT	1934
QY	1441	GTAGCGTGGCTGCAGTCTTCTGTCCCTGTGCTGTCATGTGGCTCACTCACTCACTC	1500

Db	1935	GTAGCGCTGGCTGCAGTCTTGTGCTGCTGGTCATGGCTCCTCACTACTCTCATCTC	1994
QY	1501	ATATTAGTGTGCTTGGCACTGGAG	1526
Db	1995	ATATTAGTGTGCTTGGCACTGGAG	2020
RESULT 7			
LOCUS	AX405795	1907 bp	DNA linear PAT 14-JUN-2002
DEFINITION	Sequence 210 from Patent WO0222660.		
ACCESSION	AX405795		
VERSION	AX405795.1	GI:21439029	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F., Xue, A.J., Zhou, Y., Zhehrman, T. and Drmanac, R.T.		
TITLE	Novel nucleic acids and polypeptides		
JOURNAL	Patent: WO 0222660-A 210 21-MAR-2002;		
FEATURES	HYSEQ, INC. (US)		
source	Location/Qualifiers		
	1..1907		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
CDS	101..1258		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAD34902.1"		
	/db_xref="GI:21439030"		
	/translation="MPEFLLLLVLLLLLEDACGAQGDGGHTVLGPESGTLTSINYP QTPNSTVCWEIRVRKMGKRVIRKFGDFIEDSDSCHFNLYRIYNGIGVSRTEIGKYK GLGLQNMHSIESKNGNITLFFMSGIVSGRGLASVYDKODLITCLDSATNFLEPE FSYCPAGCLLPPAEISGTPICPHGRSDSPCLAGVHAGVSVNTLGGQISVTSKGTIPY YESSLANVTVSVGHLSLSTLFPFKSCYGTGLMGESGDRGSNNSTVLEWTDHTGO ENSWKPKSOAEKWTALGAFADDEVOMLOIDLNEKKITGITGTGSTMVSTITWCLP TESCVMGMGRNGLCTESLVVEDDKIFQGNKRIITRVRNFFLPIIARLU"		
BASE COUNT	548 a 449 g 504 t		
ORIGIN			
Query Match	66.3%;	Score 1451.8;	DB 6; Length 1907;
Best Local Similarity	99.0%;	Pred. No. 0;	
Matches 1577;	Conservative	0; Mismatches	2; Indels 14; Gaps 11;
QY	1	ATGCCTCTGTTCCTCCTCTTACTTGTCTCTCTGCTGCTGCTGCTGAGGACGCTGGAGCC	60
Db	101	ATGCCTCTGTTCCTCCTCTTACTTGTCTCTCTGCTGCTGAGGACGCTGGAGCC	160
QY	61	CAGCAAGGTGATGGATGTGACACACTGTACTAGGCCTCCTGAGAGTGGAACTTACATCC	120
Db	161	CAGCAAGGTGATGGATGTGACACACTGTACTAGGCCTCCTGAGAGTGGAACTTACATCC	220
QY	121	ATAAATACCACAGACCTATCCCAACAGACACTGTTTGTGAAATGGGAGATCCGCTGTAAG	180
Db	221	ATAAATACCACAGACCTATCCCAACAGACACTGTTTGTGAAATGGGAGATCCGCTGTAAG	280
QY	181	ATGGGAGAGAGATTCCGATCAAAATTTGGTGACTTTGACATTTGAAGATTCTGATTTCTGT	240
Db	281	ATGGGAGAGAGATTCCGATCAAAATTTGGTGACTTTGACATTTGAAGATTCTGATTTCTGT	340
QY	241	CACCTTAATTACTTGAGATTTTAAATGGAATTTGGAGTGCAGCAGAACTGAAATAGGCAAA	300
Db	341	CACCTTAATTACTTGAGATTTTAAATGGAATTTGGAGTGCAGCAGAACTGAAATAGGCAAA	400
QY	301	TACTGTGGTCTGGGGTTCGAAATGAACCATTCATTAATGAATCAAAAGGCAATGAATCACA	360
Db	401	TACTGTGGTCTGGGGTTCGAAATGAACCATTCATTAATGAATCAAAAGGCAATGAATCACA	460
QY	361	TTGCTGTTTCATGAGTGGAAATCCATGTTTCTGGACGCGGATTTTGGCCCTCATCTCTGTT	420

[illegible]

Db 1536 ACCAAGATGTAGCGCTGGCTGCAGTTCCTTGCTGCTGGTCATGCGTCCCTCACTACT 1595

Qy 1492 CTCATTCTCATATTAGTAGTGTGCTTGGCACATGGAGAACAACAGAAAGAAAAAAGCAGAAGC 1551
|||||
Db 1596 CTCATTCTCATATTAGTAGTGTGCTTGGCACATGGAGAACAACAGAAAGAAAAAAGCAGAAGC 1655
|||||

Qy 1552 ACCTATGACTTTACCTTACTTGGACCGGGCAGGT 1584
|||||
Db 1656 ACCTATGACTTTACCTTACTTGGACCGGGCAGGT 1688
|||||

RESULT 8
HUMCUB1
LOCUS HUMCUB1 1388 bp mRNA linear PRI 07-FEB-1999
DEFINITION Human mRNA for unknown product, partial cds.
ACCESSION D29810
VERSION D29810.1 GI:704440
KEYWORDS
SOURCE Homo sapiens cell line L121 cDNA to mRNA, clone cub-1.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1388)
unpublished
2 (bases 1 to 1388)
unpublished
Shibata,T.
Submitted (08-Apr-1994) Tatsuhiko Shibata, National Cancer Center
Reserch Institute, Pathology, 1-1 Tsukiji 5 chome chuoku, Tokyo,
Tokyo 104, Japan (tel:03-3542-2511(ex.4208); Fax:03-3248-2737),
On Mar 10, 1995 this sequence version replaced gi:474985.
Submitted (08-Apr-1994) to DBJ by:
Tatsuhiko Shibata
National Cancer Center Reserch Institute
1-1 Tsukiji 5 chome
Chuoku, Tokyo 104
Japan
Phone: 03-3542-2511 x4208
Fax: 03-3248-2737.

FEATURES
source Location/Qualifiers
1..1388
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Li21"
<1..1097
/codon_start=3
/product="unknown"
/protein_id="BAA18909.1"
/db_xref="GI:704441"
/translation="DGGCHTVLGPESGTLSINSIPQYPNSTVCEWIRVMKGERVIRI
KFDFIEDSDSCHFNFLRYINGVSRTEIGKYGLGLQINHSIEKGNEITLLFFMS
GIHVSGRGFLASYSVIDQRNSNYLFHCIOFFWNLSVSAPAGCLLPFAEISTGFPH
GYRDSPLCMAGVHAGVSNLTGGQISVIISKGIYPYESSLANNVTSVGHLILQVF
TFKTSGCYGTLMGESGDGRSSNNSITVLEWTDDHTGOENSMKKPKSOAEKWTALGA
ATDEYQLWDLNKEKITGIITFTGSMTVSTIMCLPTESCITYMGNGLCTSLVEVE
QDKIFQGNKRITLRNRNFPPFIARLE"

CDS
BASE COUNT 432 a 289 c 302 g 364 t 1 others
ORIGIN

Query Match 52.9%; Score 1158.4; DB 9; Length 1388;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1358; Conservative 0; Mismatches 12; Indels 23; Gaps 17;

Qy 68 GTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCTTACATCCATAAACT 127
Db 1 GTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCTTACATCCATAAACT 60

Qy 128 ACCCAGACCTATCCCACACACTGTTTGTGAATGGGAGATCGTGTAAGATGGGAG 187
Db 61 ACCCAGACCTATCCCACACACTGTTTGTGAATGGGAGATCGTGTAAGATGGGAG 120

QY 188 AGAGAGTTGGCATCAAAATTTGGTGACTTTGACATTTGAAGATTCTGATTTCTTGTCACCTTTA 247
|||||
Db 121 AGAGAGTTGCATCAAAATTTGGTGACTTTGACATTTGAAGATTCTGATTTCTTGTCACCTTTA 180
QY 248 ATTACTTGAGATTTTATAATGAATTTGGAGTGCAGCAGAACTGAAATAGGCAAAATAGTGTG 307
|||||
Db 181 ATTACTTGAGATTTTATAATGAATTTGGAGTGCAGCAGAACTGAAATAGGCAAAATAGTGTG 240
QY 308 GTCTGGGGTGGCAATGAACCATTTCAATTTGAATCAAAAGGCAATGAAATACATTTGCTGT 367
|||||
Db 241 GTCTGGGGTGGCAATGAACCATTTCAATTTGAATCAAAAGGCAATGAAATACATTTGCTGT 300
QY 368 TCATGAGTGGAAATCCATGTTTCTGACGCGGATTTTGGCCCTCACTACTCTGTATTAGATA 427
|||||
Db 301 TCATGAGTGGAAATCCATGTTTCTGACGCGGATTTTGGCCCTCACTACTCTGTATTAGATA 360
QY 428 AAC-AAGATCTAATTAATTTGTTGGACACTGCATCCAA--TTTTTTGGAACTGAGTTTCAG 485
|||||
Db 361 AACAAAGATCTAATTAATTTGTTGGACACTGCATCCAAATTTTGGAACTGAGTTTCAG 420
QY 486 TAAGTACTG-CCACGTGTTGCTGCTTCCTTTTGGCTGAGATATCTGGAACAATTCCTC 544
|||||
Db 421 TAAGTACTGCCCCAGCTGTTGCTGCTTCCTTTTGGCTGAGATATCTGGAACAATTCCTC 480
QY 545 ATGGATATAGAGATTCCTCGCCATTTGTGCATGGCTGGTGCATGCAGGAGTAGTGCAA 604
|||||
Db 481 ATGGATATAGAGATTCCTCGCCATTTGTGCATGGCTGGTGCATGCAGGAGTAGTGCAA 540
QY 605 ACACGTTGGGGCGCCAAATCACTGTTGTAATTTAGTAAGGTTATCCCTATTATGAAGTT 664
Db 541 ACACGTTGGGGCGCCAAATCACTGTTGTAATTTAGTAAGGTTATCCCTATTATGAAGTT 600
QY 665 CTTTGGCTAACACGTCACATCTGTTGGTGGACACTT--ATCTACAAGTC-TTTTACAT 721
|||||
Db 601 CTTTGGCTAACACGTCACATCTGTTGGTGGACACTTGTATCTACAAGTC-TTTTACAT 660
QY 722 TTAAGACAAGTGGATTTATGGAACACTGGGATGGAGTCTGGT-GTATCGCGGATCCT 780
Db 661 TTAAGACAAGTGGATTTATGGAACACTGGGATGGAGTCTGGTGGTGTGATCGCGATCCT 720
QY 781 CAAATTAACAGCATCTCTGCTGAGTGGAGTGCACACACAGGCGCAAGAACAGTGG 840
|||||
Db 721 CAAATTAACAGCATCA-CTGTGCTGGAGTGGAGTGCACACACAGGCGCAAGAACAGTGG 779
QY 841 AAACCC-AAAAAGCAGCTGAAAAACCTGGACCGCTTGGGCTGCTTTTGGCCACTGA 899
|||||
Db 780 AAACCCAAAAAAGCCAGCTGAAAAACCTGGACCGCTTGGG-TGCTTTGGCCACTGA 838
QY 900 TGAATACAGTGGTTACAAATAGATTTGAATAAGGAAAAAGAAATTAACAGGCATTATAAC 959
|||||
Db 839 TGAATACAGTGGTTACAAATAGATTTGAATAAGGAAAAAGAAATTAACAGGCATTATAAC 898
QY 960 CACTGGATCCACCATGGT-GAGCAAAATACATATGTTGTCTGCTACAGAAATCTGTACAG 1019
Db 899 CACTGGATCCACCATGGT-GAGCAAAATACATATGTTGTCTGCTACAGAAATCTGTACAG 957
QY 1020 TGATGATGGCGAGAATGGAGTGTGTACAGAGAGCGCTGGT-GTGAGCAAGATAGATAT 1078
|||||
Db 958 TGATGATGGCGAGAATGGAGTGTGTACAGAGAGCGCTGGTGGTGGCAAGATAGATAT 1017
QY 1079 TTCAGGAAACAAA--GATTAATCACAGGAT-GTGGTGAATTAACATTTTGGCCACCAATTA 1135
|||||
Db 1018 TTCAGGAAACAAAAGAAATTAACAGGATGGTGGTGAATTAACATTTTGGCCACCAATTA 1077
QY 1136 TTGACGCTTTATTAGATGATCTACCCCAATGGCAGCAGAAAAATGCCATGAATGG 1195
Db 1078 TTGACGCTTTATTAGATGATCTACCCCAATGGCAGCAGAAAAATGCCATGAATGG 1136
QY 1196 AGCTGCTGGAGTGTGAGTTTATCTCTAAAGGTGCTCTCCAAAACCTTACTCAACCTCCAC 1255
|||||
Db 1137 AGCTGCTGGAGTGTGAGTTTATCTCTAAAGGTGCTCTCCAAAACCTTACTCAACCTCCAC 1196
QY 1256 CTCCTCGGAACGAATGACCTCAAAAACACTACAGCCCTCCAAAAA---TAGCCAAAG 1312

Db 1197 CTCCTCGGAACAGCAATGACCTC-AAACACTACAGCCCTCCAAAAATTTTAGCCAAAG 1255
QY 1313 GTCTGCCCCCAAAATTTAGCAACCACTAAACCTCGCAGTAGCAATGAATTTCTGTGCAC 1372
|||||
Db 1256 GTCTGCCCCCAAAATTTAGCAACCACTAAACCTCGCAGTAGCAATGAATTTCTGTGCAC 1315
QY 1373 AGACAGACAAACAACTGCCAGTCTCTGATATCAGAAATACTA---CCGTAACCTCCAAATG 1429
|||||
Db 1316 AGACAGACAAACAACTGCCAGTCTCTGATATCAGAAATTTACCTACCGTAACCTCCAAATG 1375
QY 1430 TAACCAAGATGT 1442
|||||
Db 1376 TAACCAAGATGT 1388
RESULT 9
AC013497 Homo sapiens chromosome 3 clone RP11-319J24 map 3, WORKING DRAFT
LOCUS AC013497
DEFINITION SEQUENCE, 11 unordered pieces.
AC013497
ACCESSION AC013497
VERSION AC013497.4 GI:7329299
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 192247)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 3, clone RP11-319J24
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 192247)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Bouckhealter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearlano, K., Dewar, K., Domino, M., Donlan, L., Doyle, M.,
Farrel, P., FitzHugh, W., Forrest, C., Fupke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehocky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:692347.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1852
Center clone name: 319_J_24
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182870 bases at least Q40
Consensus quality: 187032 bases at least Q30
Consensus quality: 188925 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 191247; sum-of-contigs
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-contigs

[illegible]


```

RESULT 12
LOCUS BC007117 586 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:4333276, mRNA, partial cds.
ACCESSION BC007117
VERSION BC007117.1 GI:13938000
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 586)
Direct Submission
Submitted (30-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 21 Row: n Column: 22.
Location/Qualifiers
1..586
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4333276"
/tissue_type="Skin, melanotic melanoma, high MDR."
/clone_lib="NHL_MGC_62"
/lab_host="DH10B"
/notes="Vector: pDNR-LIB"
<1..106
/codon_start=2
/product="Unknown (protein for IMAGE:4333276)"
/protein_id="AAH07117.1"
/db_xref="GI:13938001"
/translation="LLLVLLLLLEDAGAQVWEARILTAGCAEGHALS"
BASE COUNT 190 a 106 c 127 g 163 t
ORIGIN
Query Match 16.9%; Score 370.4; DB 9; Length 586;
Best Local Similarity 97.2%; Pred. No. 7.5e-98;
Matches 377; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 65 AAGGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAACCCCTTACATCCATAA 124
|||||
Db 113 AAGGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAACCCCTTACATCCATAA 172
QY 125 ACTACCCAGACACCTATCCCAACAGACACTGTTGTGTAATGGAGATCCCGTGAAGATGG 184
|||||
Db 173 ACTACCCAGACACCTATCCCAACAGACACTGTTGTGTAATGGAGATCCCGTGAAGATGG 232
QY 185 GAGAGAGATTCCCATCAATTTGGTGACTTTGACATTGACATTGACATTCTGTGCACT 244
|||||
Db 233 GAGAGAGATTCCCATCAATTTGGTGACTTTGACATTGACATTGACATTCTGTGCACT 292
QY 245 TTAATTACTTGAGATTTTATATGAATGGAGTTCAGCAGAACTGAAATAGCAAACTACT 304
|||||
Db 293 TTAATTACTTGAGATTTTATATGAATGGAGTTCAGCAGAACTGAAATAGCAAACTACT 352
QY 305 GTGGTCTGGGGTTGCAAAATGAACCACTTCAATTGAATCAAAAGCAATGAATCACAATTGC 364

```

```

Db 353 GTGCTGGGGTTCAAATGAACCAATCAATTGAATCAAAAGCAATGAATCAATGC 412
|||||
QY 365 TGTTCATGAGTGAATCATGTTCTTGACGCGGATTTTGGCCTCATACACTCTGTATAG 424
|||||
Db 413 TGTTCATGAGTGAATCATGTTCTTGACGCGGATTTTGGCCTCATACACTCTGTATAG 472
|||||
QY 425 ATAAACAAGATCTAATTACTTGTGTTGGA 452
|||||
Db 473 ATAAACAAGGTAATTTTCACCTTTTGCA 500
|||||

RESULT 13
LOCUS AC126975/c
DEFINITION Rattus norvegicus clone CH230-234N3, *** SEQUENCING IN PROGRESS
***, 74 unordered pieces.
ACCESSION AC126975
VERSION AC126975.1 GI:21731383
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 203102)
REFERENCE
AUTHORS
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, X., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Hui, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlssoon, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savory, G.,
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sison, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 203102)
REFERENCE
AUTHORS
Worley, K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

COMMENT

Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GJ2Q
 Center clone name: CH230-234N3
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 145293 bases at least Q40
 Consensus quality: 154399 bases at least Q30
 Consensus quality: 161802 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 74 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1
 * 1310: contig of 1310 bp in length
 * 1311: gap of unknown length
 * 1411: contig of 1087 bp in length
 * 2498: gap of unknown length
 * 2598: gap of unknown length
 * 3718: contig of 1121 bp in length
 * 3819: gap of unknown length
 * 4974: contig of 1156 bp in length
 * 5074: gap of unknown length
 * 6118: contig of 1044 bp in length
 * 6219: gap of unknown length
 * 7275: contig of 1058 bp in length
 * 7377: gap of unknown length
 * 8394: contig of 1018 bp in length
 * 8494: gap of unknown length
 * 8999: contig of 1405 bp in length
 * 9999: gap of unknown length
 * 10000: contig of 1011 bp in length
 * 11010: contig of 1011 bp in length
 * 11011: gap of unknown length
 * 11111: contig of 1319 bp in length
 * 12430: gap of unknown length
 * 12530: contig of 1704 bp in length
 * 14233: contig of 1704 bp in length
 * 14334: gap of unknown length
 * 14334: gap of unknown length
 * 15599: contig of 1266 bp in length
 * 15600: gap of unknown length
 * 15699: gap of unknown length
 * 16728: contig of 1027 bp in length
 * 16727: gap of unknown length
 * 16827: contig of 2092 bp in length
 * 18918: gap of unknown length
 * 19019: contig of 1819 bp in length
 * 20837: gap of unknown length
 * 20937: gap of unknown length
 * 20938: contig of 2099 bp in length
 * 23037: gap of unknown length
 * 23137: contig of 1754 bp in length
 * 24890: gap of unknown length
 * 24991: gap of unknown length
 * 26683: contig of 1693 bp in length
 * 26684: gap of unknown length
 * 26783: gap of unknown length
 * 28087: contig of 1304 bp in length
 * 28088: gap of unknown length
 * 28187: contig of 1660 bp in length
 * 29847: gap of unknown length
 * 29848: contig of 1218 bp in length
 * 31165: gap of unknown length
 * 31166: gap of unknown length
 * 31266: contig of 1062 bp in length
 * 32328: gap of unknown length
 * 32427: contig of 1127 bp in length
 * 33554: gap of unknown length
 * 33555: gap of unknown length

* 33655: contig of 1040 bp in length
 * 34595: gap of unknown length
 * 34795: contig of 1713 bp in length
 * 36507: gap of unknown length
 * 36807: gap of unknown length
 * 38322: contig of 1715 bp in length
 * 38422: gap of unknown length
 * 38423: contig of 1974 bp in length
 * 40397: gap of unknown length
 * 40497: contig of 2199 bp in length
 * 42695: gap of unknown length
 * 42696: gap of unknown length
 * 42795: gap of unknown length
 * 43853: contig of 1058 bp in length
 * 43854: gap of unknown length
 * 43954: contig of 1055 bp in length
 * 45009: gap of unknown length
 * 45108: gap of unknown length
 * 47026: contig of 1918 bp in length
 * 47126: gap of unknown length
 * 47127: contig of 2069 bp in length
 * 49195: gap of unknown length
 * 49295: gap of unknown length
 * 49396: contig of 2453 bp in length
 * 51749: gap of unknown length
 * 51848: gap of unknown length
 * 54202: contig of 2354 bp in length
 * 54302: gap of unknown length
 * 56606: contig of 2304 bp in length
 * 56706: gap of unknown length
 * 56607: gap of unknown length
 * 58300: contig of 1594 bp in length
 * 58301: gap of unknown length
 * 58400: gap of unknown length
 * 59928: contig of 1528 bp in length
 * 60028: gap of unknown length
 * 62029: contig of 2199 bp in length
 * 62227: gap of unknown length
 * 62327: gap of unknown length
 * 64445: contig of 2118 bp in length
 * 64545: gap of unknown length
 * 66642: contig of 2097 bp in length
 * 66742: gap of unknown length
 * 68545: contig of 1803 bp in length
 * 68645: gap of unknown length
 * 70370: contig of 1725 bp in length
 * 70470: gap of unknown length
 * 72796: contig of 2326 bp in length
 * 72896: gap of unknown length
 * 75152: contig of 2256 bp in length
 * 75252: gap of unknown length
 * 75253: gap of unknown length
 * 77649: contig of 2397 bp in length
 * 77749: gap of unknown length
 * 80695: contig of 2946 bp in length
 * 80795: gap of unknown length
 * 83796: contig of 3002 bp in length
 * 83797: gap of unknown length
 * 83798: gap of unknown length
 * 86763: contig of 2866 bp in length
 * 86863: gap of unknown length
 * 88992: contig of 3029 bp in length
 * 89992: gap of unknown length
 * 92105: contig of 2113 bp in length
 * 92205: gap of unknown length
 * 94330: contig of 2125 bp in length
 * 94430: gap of unknown length
 * 96835: contig of 2405 bp in length
 * 96935: gap of unknown length
 * 99553: contig of 2618 bp in length
 * 99554: gap of unknown length
 * 102601: contig of 2948 bp in length
 * 102701: gap of unknown length
 * 103356: contig of 2655 bp in length
 * 105456: gap of unknown length
 * 105457: contig of 3978 bp in length
 * 109434: gap of unknown length
 * 109534: gap of unknown length
 * 112973: contig of 3439 bp in length
 * 113073: gap of unknown length

Query Match
 Best Local Similarity
 Matches 407; Conservative

16.2%; Score 353.8; DB 2; Length 203102;
 85.3%; Pred. No. 1.2e-92;
 0; Mismatches 67; Indels 3; Gaps 1;

```
QY 1714 TCTGCAGAGTATGCTACGCCACTGTAGGAGGAATGTTGGTACACTTTCATCAAGATCT 1773
|||||
Db 180181 TCCACAGAAATAGCGACAGCCCTCTGCGGAGGAATGTTGGCACACTCCATCAGAGATCT 180122
QY 1774 ACCTTTAACACAGAGAGGAAGAAAGACAGCGCTATGAGACCTAGATGCTTACAACTCA 1833
|||||
Db 180121 ACCTTTAACCTGAAGAGGAAGAAAGACGAGCTACGAGATCTAGACCCCTTACAACTCA 180062
QY 1834 CCAGGCGAGGAGTTTATCATGCTCTATGCTGAACACACTCCCAATATACGGGGCTGAGTAT 1893
|||||
Db 180061 CCAGTACAGGAAGTGTATCATGCTTACGCTGAGCGCTGCGCGTACACGGGGCTGAGTAT 180002
QY 1894 GCAACCCCAATCATCATGAGATGTCAGGAGCACCCACAACTTCACTAGTGGTCCAGCCCTCC 1953
|||||
Db 180001 GCAACCCCAATCGTATGAGATGTCAGGAGCACCTCCAGACGCTCAGTGGTGGTCCCTCC 179942
QY 1954 ACATCCACTTTCAGGCTACGGGGAACCAACTCCCTCCACTAGTGGGAACTTACAACTACA 2013
|||||
Db 179941 ACATCCACTTTCAGAACTGCAAGGAACCAAGCTCCCTCCGATAGTGGGAACTTACAACTACA 179882
QY 2014 CTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGCCAGCCAGTATGATACCCCGAAAGCT 2073
|||||
Db 179881 CTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGCCAGCTCAGTACGACACCCCAAGCT 179822
QY 2074 GGGAGCCAGGCTTACCTGCCCCAGACGAATGTTGTACCAAGTCCACAGACGACACAA 2133
|||||
Db 179821 GGGAGCCAG---CAGCTGCCCCAGACGAATGTTGTACCAAGTCCGCGAGGACCCAG 179765
QY 2134 GAAGTATCAGGACGAGGAGGATGGGAATGATGATGTTTAAAGAAATCCTTTGA 2190
|||||
Db 179764 GAAGCATCAGGAGGAGGAGGATGAGAAATTTGATGCTTTTAAAGAAACCCCTTTGA 179708

RESULT 14
AC091213 Homo sapiens chromosome 3 clone CTD-2011L5, linear HTG 24-JUN-2002
LOCUS 8 unordered pieces.
DEFINITION AC091213
ACCESSION AC091213
VERSION AC091213.13 GI:21539673
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164216)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burtell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Huiy,J., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsone,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
```

FEATURES

```
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Tanton,A., Svatok,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tutson,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

2 (bases 1 to 164216)
Worley,K.C.
Direct Submission
Submitted (04-APR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 164216)
Worley,K.C.
Direct Submission
Submitted (24-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 23, 2002 this sequence version replaced gi:21539104.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCIB
Center clone name: CTD-2011L5
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 161549 bases at least Q40
Consensus quality: 162327 bases at least Q30
Consensus quality: 162604 bases at least Q20
Estimated insert size: 157933; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3452: contig of 3452 bp in length
* 3453 3552: gap of unknown length
* 3553 10549: contig of 6997 bp in length
* 10550 10649: gap of unknown length
* 10650 26000: contig of 15351 bp in length
* 26001 26100: gap of unknown length
* 26101 44000: contig of 17900 bp in length
* 44001 44100: gap of unknown length
* 44101 58020: contig of 13920 bp in length
* 58021 58120: gap of unknown length
* 58121 79695: contig of 21575 bp in length
* 79696 79795: gap of unknown length
* 79796 119936: contig of 40141 bp in length
* 119937 120036: gap of unknown length
* 120037 164216: contig of 44180 bp in length.
* Location/Qualifiers
```

```
source 1..164216
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/cloned="C1201115"
BASE COUNT 52340 a 31466 c 30520 g 49178 t 712 others
ORIGIN
Query Match 10.68; Score 232.4; DB 2; Length 164216;
Best Local Similarity 99.68; Pred. No. 1e-56;
Matches 233; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 65 AAGTGTGATGGTGGACACACTGTACTAGGCGCTGAGAGTGGAAACCCCTTACATCCATAA 124
Dbb 31855 AAGGTGTGATGGACACACTGTACTAGGCGCTGAGAGTGGAAACCCCTTACATCCATAA 31914

Qy 125 ACTACCCACAGACCTATCCCAACAGCACTGTTGTGAATGGGAGATCCGTTAAAGATGG 184
Dbb 31915 ACTACCCACAGACCTATCCCAACAGCACTGTTGTGAATGGGAGATCCGTTAAAGATGG 31974

Qy 185 GAGAGAGATTCGCATCAAAATTTGGTGACTTTGACATTTGAAGATTTCTGATTTCTTGCTACT 244
Dbb 31975 GAGAGAGATTCGCATCAAAATTTGGTGACTTTGACATTTGAAGATTTCTGATTTCTTGCTACT 32034

Qy 245 TTAATTACTTGAGAAATTTAATGGAATTTGGAGTCAGCAGAACTGAAATAGGCA 298
Dbb 32035 TTAATTACTTGAGAAATTTAATGGAATTTGGAGTCAGCAGAACTGAAATAGGTA 32088

RESULT 15
LOCUS AX118820 1620 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 5 from Patent WO0129219.
ACCESSION AX118820
VERSION AX118820.1 GI:14035776
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1620)
AUTHORS Donoho, G., Turner, C.A., Nehls, M., Friedrich, G., Zambrowicz, B. and
Sands, A.T.
TITLE Human cub-domain-containing protein and gene encoding the same
JOURNAL Patent: WO 0129219-A 5 26-APR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source 1..1620
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 456 a 353 c 422 g 389 t
ORIGIN
Query Match 10.2%; Score 224.2; DB 6; Length 1620;
Best Local Similarity 52.0%; Pred. No. 1.6e-54;
Matches 607; Conservative 0; Mismatches 518; Indels 42; Gaps 3;

Qy 62 AGCAAGGTGATGGTGGACACACTGTACTAGGCGCTGAGAGTGGAAACCCCTTACATCCA 121
Dbb 107 AGCTGGGTGATGGTGGACACACTGTACTAGGCGCTGAGAGTGGAAACCCCTTACATCCA 166

Qy 122 TAACTACCCACAGACCTATCCCAACAGCACTGTTGTGAATGGGAGATCCGTTGTAAGA 181
Dbb 167 AGAATTATCCCGGACCTACCCCAATCACACTGTTTGGAAAGACAATTTACAGTACCAA 226

Qy 182 TGGGAGAGAGATTCGCATCAAAATTTGGTGACTTTGACATTTGAAGATTTCTGATTTCTGTC 241
Dbb 227 AGGGGAAAGAGACTGATTTCTGAGTTGGGAGATTTGGATATCGAA 270

Qy 242 ACTTTAATTACTTGAGAAATTTAATGGAATTTGGAGTCAGCAGAACTGAAATAGGCAAT 301
Dbb 271 --TCCAGACCTGTGCTGCTGACTATCTTCTTCCACGACCTCTTCAGATCAATATGCTC 328
```

Search completed: January 21, 2003, 07:13:14
Job time : 6306.73 sec

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 13:50:07 ; Search time 29.4488 Seconds
(without alignments)
2488.651 Million cell updates/sec

Title: US-10-060-830-6

Perfect score: 90

Sequence: 1 caacttcagttggtcagcccc.....tttgaagatgatgtgcttt 275

Scoring table:

OLIGO Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1687582

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cqn2.1/USPTO_spool/US10060830/runat_16012003_092701_1480/app-query.fasta_1.1109
-DB=A.Geneseq_101002 -OFWT=fastan -SUFFIX=olin2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830_SCGN_1.1.36 @runat_16012003_092701_1480 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEOURRY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

A.Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	94.4	197	22	ABG01298
2	85	94.4	669	21	AAV0539
3	9	10.0	73	22	AAU62673
4	9	10.1	505	23	AAU83606
5	9	10.1	505	23	AAE14447
6	8	8.9	283	20	AAI35006
7	8	9.0	440	20	AAI37391
8	8	9.0	471	22	AAU60046
9	8	8.9	638	20	AAW86014
10	8	8.9	685	22	AAU07861
11	8	8.9	971	22	AAU62469
12	8	8.9	1162	22	ABW62516
13	7	7.9	31	21	AAW38576
14	7	7.9	36	22	ABG18047
15	7	7.9	37	19	AAW75199
16	7	7.8	49	22	ABG15514
17	7	7.8	57	22	ABG25915
18	7	7.8	57	22	AAU14916
19	7	7.9	66	22	AAU41021
20	7	7.8	71	22	AAU14829
21	7	7.9	76	22	AAW82562
22	7	7.9	80	22	AAU48986
23	7	7.8	82	22	AAU19770
24	7	7.8	82	23	ABP47990
25	7	7.8	86	22	AAU22930
26	7	7.9	88	22	AAW81286
27	7	7.9	88	22	AAW81237
28	7	7.9	90	22	AAO2815
29	7	7.9	97	17	AAW01110
30	7	7.9	98	21	AAW28183
31	7	7.9	102	22	AAU29538
32	7	7.8	102	22	ABP06512
33	7	7.9	110	20	AAW86123
34	7	7.8	112	22	AAW85215
35	7	7.8	119	21	AAW55699
36	7	7.8	119	21	AAW59587
37	7	7.8	119	22	AAU33028
38	7	7.8	129	21	AAW55698
39	7	7.8	129	21	AAW55698
40	7	7.9	138	22	AAU50353
41	7	7.8	145	22	ABG01319
42	7	7.8	145	22	AAU27659
43	7	7.8	151	21	AAW25232
44	7	7.9	153	22	AAW58577
45	7	7.8	154	22	AAU40344

ALIGNMENTS

RESULT 1

ABG01298
ID ABG01298 standard; Protein; 197 AA.

XX ABG01298;

XX 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #1289.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

PD

PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR N-PSDB; AAS65485.
XX
DR New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 31657; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 197 AA:

Alignment Scores:
Pred. No.: 1.08e-73 Length: 197
Score: 85.00 Matches: 85
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 94.44% Indels: 0
DB: 22 Gaps: 0

US-10-060-830-6 (1-275) x ABG01298 (1-197)

QY 3 ACTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAACCTCCCCA 62
Db 113 ThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProPro 132
QY 63 CTAGTGGGAACCTTACAATACACTTCTCCAGGACTGACAGCTGCTCTCAGCCAGGCC 122
Db 133 LeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAla 152
QY 123 CAGTATGATACCCCGAAGCTGGGAGCCAGGTCTACCTGCCCCAGACGAATTGGTGTAC 182
Db 153 GlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTyr 172
QY 183 CAGGTGCCACAGACGACACAAAGATATCAGGAGCAGGAGGATGGGAATGTGATCTT 242
Db 173 GlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGlyGluCysAspVal 192
QY 243 TTTAAAGAAATCCTT 257
Db 193 PheLysGluLeuLeu 197

RESULT 2

AAY70539
ID AAY70539 standard; Protein; 669 AA.
XX
AC AAY70539;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human Factor 8 Homologue.
XX
KW Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW cerebroprotective; therapeutic; coagulation related disorder;
KW haemophilia; stroke; screening.
XX
OS Homo sapiens.
XX
PN WO200012532-A1.
XX
PD 09-MAR-2000.
XX
PF 20-AUG-1999; 99WO-US19047.
XX
PR 31-AUG-1998; 98US-0098521.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Rosteck PRJ, Su W, Li XM;
XX
DR WPI: 2000-256580/22.
DR N-PSDB; AAZ51872.
XX
Factor 8 homolog polypeptides and nucleic acids encoding them for
treating coagulation related disorders such as hemophilia and stroke -
Claim 3; Page 64-66; 68pp; English.
XX
The present sequence is a human Factor 8 homologue (F8H),
a coagulation cofactor which is selectively expressed in
CC haematopoietic, heart and reproductive tissues. It has haemostatic and
CC cerebroprotective activities. The F8H contains a Factor 5/8 signature
CC and is useful as a therapeutic for treating coagulation related diseases
CC such as haemophilia and stroke. The nucleic acid is useful as
CC hybridisation probe and amplification primer for detecting deficiencies
CC in the level of F8H mRNA, for screening F8H gene mutations and for
CC monitoring regulation of gene expression. Fragments of the nucleic acid
CC are also useful as diagnostic probes and primers, and can be used in
CC screening methods such as those using DNA chips. The present sequence is
CC also useful as a target to screen therapeutically useful modulators
CC of the F8H.
XX
SQ Sequence 669 AA;

Alignment Scores:

Pred. No.: 9.03e-74 Length: 669
Score: 85.00 Matches: 85
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 94.44% Indels: 0
DB: 21 Gaps: 0

US-10-060-830-6 (1-275) x AAY70539 (1-669)

QY 3 ACTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAACCTCCCCA 62
Db 595 ThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProPro 604
QY 63 CTAGTGGGAACCTTACAATACACTTCTCCAGGACTGACAGCTGCTCTCAGCCAGGCC 122
Db 605 LeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAla 624
QY 123 CAGTATGATACCCCGAAGCTGGGAGCCAGGTCTACCTGCCCCAGACGAATTGGTGTAC 182
Db 625 GlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTyr 644

QY 183 CAGTGCCACAGACACAGAAGTATCAGGAGCAGGAGGATGGGAATGTGATGTT 242
|||||
Db 645 GlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGlyGluCysAspVal 664
QY 243 TTTAAAGAAATCCTT 257
|||||
Db 665 PheLysGluIleLeu 669
RESULT 3
AAU62673
ID AAU62673 standard; Protein; 73 AA.
XX AC
XX AAU62673;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #23569.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001W0-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jèn S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59628.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 23868; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: the sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 73 AA;

Alignment Scores:
Pred. No.: 3.99 Length: 73
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.00% Indels: 0
DB: 22 Gaps: 0
US-10-060-830-6 (1-275) x AAU62673 (1-73)
QY 207 GTATCAGGACGACGAGGATGGGAA 233
|||||
Db 1 ValSerGlyAlaGlyArgAspGlyGlu 9
RESULT 4
AAU83606
ID AAU83606 standard; Protein; 505 AA.
XX AC
XX AAU83606;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human PRO protein, Seq ID No 30.
XX
KW Human; secreted protein; PRO; tumour; lung cancer; colon cancer;
KW breast cancer; prostate tumour; rectal tumour; liver tumour;
KW pericyte cell proliferation; chondrocyte cell proliferation;
KW tumour necrosis factor-alpha.
XX
OS Homo sapiens.
XX
PN WO200208288-A2.
XX
PD 31-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-US21066.
XX
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220385P.
PR 25-JUL-2000; 2000US-220605P.
PR 25-JUL-2000; 2000US-220607P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220638P.
PR 25-JUL-2000; 2000US-220664P.
PR 25-JUL-2000; 2000US-220666P.
PR 26-JUL-2000; 2000US-220893P.
PR 28-JUL-2000; 2000WO-US20710.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 15-SEP-2000; 2000US-000000P.
PR 10-NOV-2000; 2000WO-US30873.
PR 28-NOV-2000; 2000US-253846P.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001WO-US17092.
XX
XX (GETH) GENENTECH INC.
XX
XX Baker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WI;
XX
XX WPI; 2002-172001/22.
DR N-PSDB; ABK33550.
XX
PT One hundred and twenty two nucleic acids encoding PRO polypeptides,
PT useful for treating a PRO related disorder and for diagnosing tumours
PT such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal
PT tumour or liver tumour -
XX
PS Claim 11; Figure 30; 359pp; English.

XX The invention relates to one hundred and twenty two nucleic acids
 CC encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides
 CC encode human secreted proteins. The PRO nucleic acids, polypeptides,
 CC agonists and antagonists are useful for treating a PRO related disorder.
 CC The PRO polypeptides are useful for diagnosing tumours, especially lung
 CC cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or
 CC liver tumour. The PRO polypeptides are useful for stimulating the
 CC proliferation of, or gene expression, in pericyte cells, for stimulating
 CC the proliferation or differentiation of chondrocyte cells, for
 CC stimulating the release of tumour necrosis factor-alpha from human blood,
 CC for stimulating or inhibiting the proliferation of normal human dermal
 CC fibroblast cells. The PRO polypeptide may also be used as molecular
 CC weight markers and for tissue typing. The PRO nucleic acids have
 CC applications in molecular biology, including use as hybridisation probes,
 CC and in chromosome and gene mapping. AAU83592-AAU83713 represent human PRO
 CC protein sequences of the invention.

XX Sequence 505 AA:

Alignment Scores:
 Pred. No.: 2.98 Length: 505
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.11% Indels: 0
 DB: 23 Gaps: 0

US-10-060-830-6 (1-275) x AAU83606 (1-505)

QY 228 CATCCCTTCCTCCTCTGATACCTTCTT 202

Db 12 HisProPheLeuLeuLeuLeuLeu 20

RESULT 5

ID AAEL14447 standard; Protein: 505 AA.

AC AAEL14447;

DT 26-MAR-2002 (first entry)

DE Human drug metabolising enzyme (DME)-10.

XX Human; drug metabolising enzyme; DME-10; autoimmune; inflammatory;
 KW cell proliferative; developmental; endocrine; eye; metabolic; AIDS;
 KW gastrointestinal disorder; liver disorder; cancer; arteriosclerosis;
 KW adult respiratory distress syndrome; anaemia; epilepsy; hypothyroidism;
 KW hypothalamus; pituitary; diabetes; hypogonadism; conjunctivitis;
 KW glaucoma; cystic fibrosis; hypercholesterolaemia; gastritis;
 KW peptic ulcer; hepatitis; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..31

FT Protein /label= Signal_peptide

FT Domain /label= Mature_DME-10

FT /label= Transmembrane_domain

XX WO200190334-A2.

XX 29-NOV-2001.

XX 25-MAY-2001; 2001WO-US17150.

XX 25-MAY-2000; 2000US-207901P.

XX 01-JUN-2000; 2000US-208983P.

XX 07-JUN-2000; 2000US-209861P.

XX 15-JUN-2000; 2000US-211825P.

XX 22-JUN-2000; 2000US-213744P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Sanjanwala MS, Baughn MR, Gandhi AR, Ring HZ, Elliott V;

XX Walia NK, Yang J, Khan FA, Ramkumar J, Tang YT, Hafalia A, Lal P;

XX Nguyen DB, Yao MG, Lee EA, Tribouley CM, Patterson C, Lu Y;

XX Burford N, Ding L, Bruns CM, Kearney L, Reddy R;

XX WPI; 2002-097650/13.

XX N-PSDB; AAD24015.

XX New human drug metabolizing enzymes and polynucleotides encoding the
 PT enzyme for diagnosing, preventing or treating cell proliferative,
 PT autoimmune/inflammatory, endocrine, eye, metabolic and gastrointestinal
 PT disorders

XX Claim 1; Page 145-146; 158pp; English.

XX The present sequence is human drug metabolising enzyme (DME)-10.
 CC DME polypeptide, polynucleotide and modulators are useful for
 CC diagnosis, treatment and prevention of autoimmune/inflammatory,
 CC cell proliferative, developmental, endocrine, eye, metabolic,
 CC and gastrointestinal disorders, including liver disorders.

XX The autoimmune/inflammatory disorders treatable include

CC AIDS, adult respiratory distress syndrome, Addison's disease,
 CC allergies, anaemia, asthma, atherosclerosis, osteoporosis, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, Crohn's disease, atopic
 CC dermatitis, diabetic mellitus, Graves' disease, glomerulonephritis,
 CC rheumatoid arthritis, scleroderma, systemic lupus erythematosus,
 CC systemic sclerosis, ulcerative colitis, haemodialysis and uveitis,
 CC viral, bacterial, fungal, parasitic, protozoal, helminthic infections
 CC and trauma, and cell proliferative disorders such as cancer, actinic
 CC keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis,
 CC hepatitis and psoriasis. Developmental disorders include anaemia, renal
 CC tubular acidosis, epilepsy, hypothyroidism and cataract, and endocrine
 CC disorders include disorders of hypothalamus and pituitary, disorders
 CC associated with hypopituitarism, including sarcoidosis, diabetes
 CC insipidus, hypogonadism, disorders associated with hypothyroidism
 CC including goitre, acute thyroiditis, Graves' disease, disorders
 CC associated with hyperparathyroidism, pancreatic disorders such as type I
 CC or type II diabetes mellitus, disorders associated with adrenals such as
 CC hyperplasia, Cushing's disease, endometriosis, infertility,
 CC hypergonadal disorders, and gynaecomastia. Eye disorders include
 CC conjunctivitis, keratitis, glaucoma and macular degeneration, and
 CC metabolic disorders include diabetes, cystic fibrosis, goitre,
 CC hypercholesterolaemia, hypoglycaemia, hyperlipidaemia, lysosomal storage
 CC diseases, obesity, phenylketonuria and hypocalcaemia. Also the molecules
 CC are useful for treating gastrointestinal disorders such as dysphagia,
 CC gastritis, peptic ulcer, cholelithiasis, cirrhosis, hepatitis,
 CC hyperbilirubinaemia, constipation, diarrhoea, jaundice, Wilson's disease,
 CC thrombosis and hepatic tumours. The DME polypeptide is also useful for
 CC screening its agonist or antagonist.

XX Sequence 505 AA:

Alignment Scores:

Pred. No.: 2.98 Length: 505
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.11% Indels: 0
 DB: 23 Gaps: 0

US-10-060-830-6 (1-275) x AAEL14447 (1-505)

QY 228 CATCCCTTCCTCCTCTGATACCTTCTT 202

Db 12 HisProPheLeuLeuLeuLeuLeu 20

RESULT 6

AAU835006

ID AAU835006 standard; Protein: 283 AA.

XX

AC AAY35006;
 XX 13-SEP-1999 (first entry)
 XX Chlamydia pneumoniae transmembrane protein sequence.
 DE
 XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.
 XX Chlamydia pneumoniae.
 OS
 XX WO9927105-A2.
 PN
 XX 03-JUN-1999.
 PD
 XX 20-NOV-1998; 98WO-IB01890.
 PF
 XX 04-NOV-1998; 98US-0107078.
 PR
 XX 21-NOV-1997; 97FR-0014673.
 PP
 XX (GEST) GENSET.
 PA
 XX Griffais R;
 PI
 XX WPI; 1999-357842/30.
 DR
 XX Genome sequence of Chlamydia pneumoniae
 PT
 XX Page 909; Disclosure; 1912pp; English.
 PS
 XX AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 XX
 SQ Sequence 283 AA;
 Alignment Scores:
 Pred. No.: 30.2 Length: 283
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.89% Indels: 0
 DB: 20 Gaps: 0
 US-10-060-830-6 (1-275) x AAY35006 (1-283)
 QY 6 TCAGTTGGTCAGCCTCCACATCC 29
 AC
 DB 114 SerValGlyInProSerThrSer 121
 RESULT 7
 AAY37391
 ID AAY37391 standard; Protein; 440 AA.
 XX
 AC AAY37391;
 XX
 DT 07-OCT-1999 (first entry)
 DE
 XX Amino acid sequence of a Chlamydia trachomatis protein.
 KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.
 OS
 XX WO9928475-A2.
 PN
 XX 10-JUN-1999.
 PD
 XX 27-NOV-1998; 98WO-IB01939.
 PF
 XX 04-NOV-1998; 98US-0107077.
 PR
 XX 28-NOV-1997; 97FR-0015041.
 PP
 XX 17-DEC-1997; 97FR-0016034.
 PA
 XX (GEST) GENSET.
 PI
 XX Griffais R;
 DR
 XX WPI; 1999-371125/31.
 PF
 XX Genome sequence of Chlamydia trachomatis
 PS
 XX Disclosure; Page 1101; 1755pp; English.
 XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma, such as
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perihhepatitis, bartholinitis; pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.
 XX
 SQ Sequence 440 AA;
 Alignment Scores:
 Pred. No.: 28.2 Length: 440
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.99% Indels: 0
 DB: 20 Gaps: 0
 US-10-060-830-6 (1-275) x AAY37391 (1-440)
 QY 266 TCATCTTCAAAGATTCTTTAAA 243
 DB 44 SerSerSerLysAspPhePheLys 51
 RESULT 8
 AAU60046
 ID AAU60046 standard; Protein; 471 AA.
 XX
 AC AAU60046;
 XX
 DT 27-FEB-2002 (first entry)
 DE
 XX Propionibacterium acnes immunogenic protein #20942.
 DE
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX
 OS Propionibacterium acnes.
 XX
 PN WO200181581-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001WO-US12865.

XX PR 21-APR-2000; 2000US-199047P.
 XX PR 02-JUN-2000; 2000US-208841P.
 XX PR 07-JUL-2000; 2000US-216747P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX DR WPI; 2001-616774/71.
 XX DR N-PSDB; AAS59608.
 XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
 XX PT vaccinating against and diagnosing infections, especially useful for
 XX PT treating acne vulgaris.
 XX PS Example 1; SEQ ID NO 21241; 1069pp; English.
 XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 XX CC polypeptides. The proteins and their associated DNA sequences are used in
 XX CC the treatment, prevention and diagnosis of medical conditions caused by
 XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 XX CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
 XX CC P. acnes is also involved in infections of bone, joints and the central
 XX CC nervous system, however it is particularly involved in the inflammatory
 XX CC lesions associated with acne vulgaris. A method for detecting the
 XX CC presence or absence of P. acnes in a patient comprises contacting a
 XX CC sample with a binding agent that binds to the proteins of the invention
 XX CC and determining the amount of bound protein in the sample. The
 XX CC polypeptides may be used as antigens in the production of antibodies
 XX CC specific for P. acnes proteins. These antibodies can be used to
 XX CC downregulate expression and activity of P. acnes polypeptides and
 XX CC therefore treat P. acnes infections. The antibodies may also be used as
 XX CC diagnostic agents for determining P. acnes presence, for example, by
 XX CC enzyme linked immunosorbent assay (ELISA).
 XX CC Note: The sequence data for this patent did not form part of the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 471 AA;
 Alignment Scores:
 Pred. No.: 28 Length: 471
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.99% Indels: 0
 DB: 22 Gaps: 0
 US-10-060-830-6 (1-275) x AAU60046 (1-471)
 QY 180 ACACCAATCGTCGTGGGACAGTA 157
 Db 198 ThrProIleArgLeuGlyGlnVal 205
 RESULT 9
 AAU86014
 ID AAU86014 standard; Protein: 638 AA.
 XX AC AAU86014;
 XX DT 22-FEB-1999 (first entry)
 XX DE R. rubrum poly-beta-hydroxyalkanoate (PHA) synthase.
 XX KW PHA synthase; poly-beta-hydroxyalkanoate synthase; recombinant;
 XX KW transformation; transgenic plant; polymerase; 3-hydroxyacyl-ACP;
 XX KW 3-hydroxyacyl-CoA; biodegradable; PHA polyester; petrochemical;
 XX KW plastic.
 XX OS Rhodospirillum rubrum.
 XX PA Rhodospirillum rubrum.

PN US5849894-A.
 XX PD 15-DEC-1998.
 XX PF 25-NOV-1996; 96US-0756317.
 XX PR 29-NOV-1995; 95US-0007693.
 XX PR 25-NOV-1996; 96US-0756317.
 XX PA (MONS) MONSANTO CO.
 XX PI Clemente TE, Kishore GM, Mitsky TA, Stark DM;
 XX PI WPI; 1999-069811/06.
 XX DR N-PSDB; AAV80358.
 XX CC Cloned Rhodospirillum rubrum poly-beta-hydroxyalkanoate synthase
 XX PT gene - for transforming bacteria and plants to produce
 XX PT poly:hydroxy-alkanoate polyester(s)
 XX PS Disclosure; Columns 17-22; 36pp; English.
 XX CC This represents a Rhodospirillum rubrum poly-beta-hydroxyalkanoate (PHA)
 XX CC synthase. The encoding DNA is deposited under accession number ATCC
 XX CC 25903. Host cells transformed by a vector containing the PHA synthase
 XX CC encoding nucleic acid can be used for the recombinant production of the
 XX CC protein. Transformed bacteria and transgenic plants containing the DNA
 XX CC will be able to polymerise 3-hydroxyacyl-ACP and/or 3-hydroxyacyl-CoA
 XX CC substrates to produce biodegradable PHA polyesters having similar
 XX CC physical properties to petrochemical-derived plastics.
 XX SQ Sequence 638 AA;
 Alignment Scores:
 Pred. No.: 26.7 Length: 638
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.89% Indels: 0
 DB: 20 Gaps: 0
 US-10-060-830-6 (1-275) x AAU86014 (1-638)
 QY 24 ACATCCACTTTCAGGCTACGGG 47
 Db 536 ThrSerThrPheLysAlaThrGly 543
 RESULT 10
 AAU07861
 ID AAU07861 standard; Protein: 685 AA.
 XX AC AAU07861;
 XX DT 18-DEC-2001 (first entry)
 XX DE Polypeptide sequence for mammalian Spg3.
 XX KW Mammalian; reproductive-specific protein; male infertility;
 XX KW spermatogenesis; sperm count disorder; anti infertility; reproduction.
 XX OS Mammalia.
 XX PN WO200166752-A2.
 XX PD 13-SEP-2001.
 XX PF 07-MAR-2001; 2001WO-US07371.
 XX PR 07-MAR-2000; 2000US-0187518.
 XX PR 12-JAN-2001; 2001US-0261557.
 XX PA (WHEED) WHITEHEAD INST BIOMEDICAL RES.

PI Wang PJ, Page DC;
 XX WPI; 2001-570774/64.
 DR N-PSDB; AAS13625.
 XX Novel reproduction-specific protein, useful for treating disorders of
 PT reduced sperm count, enhancing/increasing sperm count and/or sperm
 PT activity -
 XX Claim 22; Fig 6; 15lpp; English.
 XX The present invention relates to the isolation of novel mammalian and
 CC human reproductive-specific proteins (AAU07859-AAU07899), and the
 CC nucleic acids encoding them. The nucleic acids encoding
 CC reproductive-specific proteins are useful for diagnosing infertility
 CC which is a result of reduced sperm count, reduced sperm motility,
 CC malformed sperm or combinations of these. The sequences of the invention
 CC are useful as markers for spermatogonial cells, for identifying genes or
 CC proteins characteristic of male infertility, diagnosing or aiding in
 CC the diagnosis of infertility in men, and for contraception in which
 CC sperm production or sperm count is reduced or defective sperm is
 CC produced. Antibodies to reproductive-specific proteins are useful for
 CC determining the presence of these proteins in a sample obtained from a
 CC man being assessed for infertility, for identifying the expression of
 CC genes in particular cell type or particular developmental stage, for
 CC studies of spermatogenesis, and for immunofluorescence of germ cells or
 CC in western blots for assessing the presence of the protein the antibody
 CC binds. The sequences of the invention are also useful for treating
 CC disorders of reduced sperm count, and for increasing sperm count and/or
 CC sperm activity. The nucleic acids of the invention are useful in gene
 CC therapy. AAU07859-AAU07882 represent the mammalian reproduction-specific
 CC proteins of the present invention.

XX Sequence 685 AA;

Alignment Scores:
 Pred. No.: 26.4 Length: 685
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.99% Indels: 0
 DB: 22 Gaps: 0

US-10-060-830-6 (1-275) x AAU07861 (1-685)

QY 226 TCCTTCCTGCTCCGATCTCT 203
 Db 618 SerLeuProAlaProAspThrSer 625

RESULT 11
 AAG62469
 ID AAG62469 standard; Protein; 971 AA.
 AC AAG62469;
 XX 10-SEP-2001 (first entry)
 DT Xylanase expression control protein.
 XX Expression control; xylanase.

OS Aspergillus oryzae.
 XX JP2001095581-A.
 PN 10-APR-2001.
 PD 29-SEP-1999; 99JP-0276484.
 XX 29-SEP-1999; 99JP-0276484.
 PR (HGET) HIGETA SHOYU KK.
 PA
 XX

DR WPI; 2001-333972/35.
 DR N-PSDB; AAH45501.
 XX New DNA combined protein for controlling expression of a xylanase gene
 PT and for efficient production of xylanase -
 XX Claim 1; Fig 2-4; 19pp; Japanese.

XX This invention relates to a protein which controls the expression of a
 CC xylanase gene. The invention includes the protein and DNA sequence, and
 CC a recombinant plasmid containing DNA encoding the xylanase expression
 CC control protein. The DNA can be used for the efficient production of
 CC xylanase. The present sequence the protein of the invention.

XX Sequence 971 AA;

Alignment Scores:
 Pred. No.: 25.1 Length: 971
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.89% Indels: 0
 DB: 22 Gaps: 0

US-10-060-830-6 (1-275) x AAG62469 (1-971)

QY 192 CAGACACACAGAGATCAGGA 215
 Db 209 GlnSerThrGlnGluValSerGly 216

RESULT 12
 ABB62516
 ID ABB62516 standard; Protein; 1162 AA.

XX ABB62516;

XX 26-MAR-2002 (first entry)
 DT Drosophila melanogaster polypeptide SEQ ID NO 14340.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.
 KW Drosophila melanogaster.
 XX WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL06619.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure; SEQ ID NO 14340; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABLO1840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 SQ Sequence 1162 AA;
 Alignment Scores:
 Pred. No.: 24.4 Length: 1162
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.89% Indels: 0
 DB: 22 Gaps: 0
 US-10-060-830-6 (1-275) x ABB62516 (1-1162)
 QY 103 GCTGCTCCTCAGCCAGCCAGT 126
 |||||
 DB 871 AlaAlaProGlnProArgProSer 878
 RESULT 13
 AAB38576
 ID AAB38576 standard; Protein: 31 AA.
 XX AC AAB38576;
 XX 30-JAN-2001 (first entry)
 XX Gene 5 human secreted protein homologous amino acid sequence #113.
 XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
 KW neotropic; neuroprotective; antibacterial; virucide; fungicide;
 KW ophthalmological; autoimmune disease; hyperproliferative disorder;
 KW cardiovascular disorder; cerebrovascular disorder; infection;
 KW wound healing; cell proliferation; chemotaxis; food additive.
 XX Mumps virus.
 XX WO200056882-A1.
 PN 28-SEP-2000.
 PD 16-MAR-2000; 2000WO-US06791.
 PF 23-MAR-1999; 99US-0125815.
 PR 10-DEC-1999; 99US-0169946.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM, Komatsoullis G;
 XX WPI; 2000-579484/54.
 DR Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition
 PT Disclosure; Page 13; 436pp; English.
 PS The polynucleotide sequences given in AAC69597-C69644 encode the human
 CC secreted proteins given in AAB38522-B38569. Sequences given in
 CC AAB38570-B38647 represent alternative proteins encoded by the genes, and
 CC also represent proteins which are homologous to the gene products. The
 CC secreted proteins have activities based on the tissues in which they are
 CC expressed. Examples of the activities include immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; neotropic; neuroprotective;
 CC antibacterial; virucide; fungicide; and ophthalmological. The proteins,
 CC polynucleotides and/or agonists/antagonists are used to prevent, treat or
 CC ameliorate medical conditions which include autoimmune diseases e.g.
 CC rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the

CC breast or liver, cardiovascular disorders e.g. cardiac arrest,
 CC cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous
 CC system disorders e.g. Alzheimer's disease, infections caused by bacteria,
 CC viruses and fungi and ocular disorders e.g. corneal infection. The
 CC polypeptides can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues and in chemotaxis. The polypeptides can also be
 CC used as a food additive or preservative to increase or decrease storage
 CC capabilities. Sequences in AAC69597-C69644 and AAB38521 are used during
 CC the isolation and characterisation of the genes of the invention.
 XX SQ Sequence 31 AA;
 Alignment Scores:
 Pred. No.: 390 Length: 31
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.87% Indels: 0
 DB: 21 Gaps: 0
 US-10-060-830-6 (1-275) x AAB38576 (1-31)
 QY 222 TTCCTGCTCCTGACTTCTT 202
 |||||
 DB 7 PheLeuLeuLeuLeuLeu 13
 RESULT 14
 AAB18047
 ID AAB18047 standard; Protein: 36 AA.
 XX AC AAB18047;
 XX 18-FEB-2002 (first entry)
 XX Novel human diagnostic protein #18038.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 XX WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS82234.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX Claim 20: SEQ ID No 48406; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences, (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 36 AA;

Alignment Scores:
 Pred. No.: 381 Length: 36
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.87% Indels: 0
 DB: 22 Gaps: 0

US-10-060-830-6 (1-275) x ABG18047 (1-36)

QY 107 GCAGCTGCAGTCCCTGGAGAG 87
 Db 18 AlaAlaValSerProGlyGlu 24

RESULT 15

AAW75199

ID AAW75199 standard; Protein; 37 AA.

XX AC AAW75199;

XX DT 29-JAN-1999 (first entry)

XX DE Human secreted protein encoded by gene 4 clone HCUHF89.

XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX OS Homo sapiens.

XX PN WO9840483-A2.

XX PD 17-SEP-1998.

XX PF 12-MAR-1998; 98WO-US04858.

XX PR 19-DEC-1997; 97US-0068368.

XX PR 14-MAR-1997; 97US-0040710.

XX PR 14-MAR-1997; 97US-0040762.

XX PR 30-MAY-1997; 97US-0048100.

XX PR 30-MAY-1997; 97US-0048189.

XX PR 30-MAY-1997; 97US-0048357.

XX PR 30-MAY-1997; 97US-0050934.

XX PR 06-JUN-1997; 97US-0048970.

XX PR 05-SEP-1997; 97US-0057765.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ferlie AM, Fischer CL, Gentz RL, Greene JM, Kyaw H;

PI Li H, Li Y, Moore PA, Rosen CA, Ruben SM, Soppet DR;

PI Wei YF, Young PE, Zeng Z;

XX

DR WPI: 1998-520811/44.

DR N-PSDB; AAV34289.

XX

PT Isolated human polynucleotide(s) encoding secretory peptide(s) -
 PT used to develop products for the diagnosis and treatment of e.g.
 PT inflammation, cancers, CNS disorders or immune system disorders

XX

PS Claim 1; Page 155; 201pp; English.

XX

CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AAV34277) for increasing the stability of the fused
 CC protein as compared to the human protein only.

CC

CC The invention relates to 28 novel genes and their fragments (nucleic
 CC acid sequences: AAV34286-V34325; amino acid sequences AAW75196-W75235)

CC

CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological

CC

CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 28

CC

CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAV34286 for described uses).

XX

SQ Sequence 37 AA;

Alignment Scores:

Pred. No.: 380 Length: 37
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.87% Indels: 0
 DB: 19 Gaps: 0

US-10-060-830-6 (1-275) x AAW75199 (1-37)

QY 210 ATACTTCTGTGTGTCTGTG 190

Db 19 IleLeuLeuValCysSerVal 25

Search completed: January 21, 2003, 14:49:03
 Job time : 31.4488 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 12:56:47 ; Search time 301.512 Seconds
(without alignments)
3222.856 Million cell updates/sec

Title: US-10-060-830-1115
Perfect score: 60
Sequence: 1 ctgctgctgagcagctgg.....gtgacacactgtactagc 60

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 16154066 seqs, 8097743376 residues

Word size : 0
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- EST:*
- 1: em_estba.*
 - 2: em_esthum.*
 - 3: em_estin.*
 - 4: em_estmu.*
 - 5: em_estcov.*
 - 6: em_estpl.*
 - 7: em_estro.*
 - 8: em_hic.*
 - 9: gb_est1.*
 - 10: gb_est2.*
 - 11: gb_hic.*
 - 12: gb_est3.*
 - 13: gb_est4.*
 - 14: gb_est5.*
 - 15: em_estfun.*
 - 16: em_estom.*
 - 17: gb_gss.*
 - 18: em_gss_hum.*
 - 19: em_gss_inv.*
 - 20: em_gss_pln.*
 - 21: em_gss_vrt.*
 - 22: em_gss_fun.*
 - 23: em_gss_mam.*
 - 24: em_gss_mus.*
 - 25: em_gss_other.*
 - 26: em_gss_pro.*
 - 27: em_gss_fod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	100.0	332	14 H80005	H80005 ys65d12.s1
2	60	100.0	388	14 H85099	H85099 ys70e05.s1
3	60	100.0	418	14 H99544	H99544 yx29g09.s1
4	60	100.0	459	9 AA85289	AA85289 al58g05.s
5	60	100.0	470	9 AI767271	AI767271 wh25f10.x
6	60	100.0	502	9 AI499115	AI499115 to05a05.x

7	60	100.0	504	9 AI565996	AI565996 tn52b03.x
8	60	100.0	526	9 AI420312	AI420312 tf06g06.x
9	60	100.0	566	12 BF434973	BF434973 7p04a07.x
10	60	100.0	605	10 AW138866	AW138866 UI-H-Bil-
11	60	100.0	693	12 BF732384	BF732384 naell1a12.
12	60	100.0	703	10 AW070902	AW070902 xa31b09.x
13	60	100.0	908	12 BG178371	BG178371 602330141
14	60	100.0	1061	14 BM905706	BM905706 AGENCOURT
15	46	76.7	692	12 BF732861	BF732861 naell6a03.
16	35	58.3	424	9 AA460989	AA460989 zx84a01.r
17	32	53.3	338	13 BI037778	BI037778 CM4-NF7029
18	32	53.3	543	12 BF691675	BF691675 602247724
19	29	48.3	534	10 AV603143	AV603143 AV603143
20	22	36.7	427	9 AJ449261	AJ449261 AJ449261
21	22	36.7	530	9 AJ448166	AJ448166 AJ448166
22	22	36.7	667	9 AL586285	AL586285 AL586285
23	19	31.7	216	13 BI285467	BI285467 UI-R-DDO-
24	18	30.0	454	13 BM430856	BM430856 1duo06A09
25	18	30.0	576	17 BH044206	BH044206 RPCI-24-3
26	18	30.0	662	13 BI232249	BI232249 RE28281.5
27	18	30.0	868	13 BG918584	BG918584 602818368
28	18	30.0	980	9 AA908991	AA908991 oll10d01.s
29	17	28.3	306	10 AV298498	AV298498 AV298498
30	17	28.3	329	13 BM086829	BM086829 498189 MA
31	17	28.3	330	13 BM032259	BM032259 497894 MA
32	17	28.3	346	9 AA318153	AA318153 EST20193
33	17	28.3	418	14 RI2756	RI2756 yf58g08.r1
34	17	28.3	462	10 BE252178	BE252178 60113772
35	17	28.3	477	10 BB858068	BB858068 BB858068
36	17	28.3	481	9 AI170345	AI170345 EST216271
37	17	28.3	488	17 AZ291045	AZ291045 RPCI-23-1
38	17	28.3	499	13 BM004612	BM004612 da334802
39	17	28.3	500	13 BI373635	BI373635 RE60977.5
40	17	28.3	516	13 BM086841	BM086841 498183 MA
41	17	28.3	527	13 BM583747	BM583747 170006872
42	17	28.3	531	12 BG134859	BG134859 EST467751
43	17	28.3	543	12 BE721709	BE721709 189317 MA
44	17	28.3	549	9 AL678252	AL678252 AL678252
45	17	28.3	561	17 AZ121328	AZ121328 RPCI-23-1

ALIGNMENTS

RESULT 1

H80005

LOCUS

DEFINITION

IMAGE:19671.3, similar to SP.BMPL_HUMAN P13497 BONE MORPHOGENETIC

PROTEIN 1 PRECURSOR ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 332)

AUTHORS

Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M., Holman

, M., Hurlman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,

Rifkin, B., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston

, R., Williamson, A., Wohldmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

TITLE

JOURNAL

COMMENT

H80005 332 bp mRNA linear EST 09-NOV-1995
ys65d12.s1 Soares retina N2b4HR Homo sapiens cDNA clone
IMAGE:19671.3, similar to SP.BMPL_HUMAN P13497 BONE MORPHOGENETIC
PROTEIN 1 PRECURSOR ;, mRNA sequence.

H80005 GI:1058094

EST.

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 332)

AUTHORS

Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M., Holman

, M., Hurlman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,

Rifkin, B., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston

, R., Williamson, A., Wohldmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

TITLE

JOURNAL

COMMENT

Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

High quality sequence stops: 298

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Insert Length: 1092 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 298.

FEATURES

source

Location/Qualifiers
1. .332
/organism="Homo sapiens"
/db_xref="GDB:3847664"
/db_xref="taxon:9606"
/clone="IMAGE:219671"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"

/note="Organ: eye; Vector: pT73D (Pharmacia) with a modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Patima Bonaldo."

BASE COUNT 85 a 67 c 89 g 88 t 3 others

ORIGIN

Query Match 100.0%; Score 60; DB 14; Length 332;
Best Local Similarity 100.0%; Pred. No. 2.7e-21;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTGCTCGAGGAGCGTGGAGCCGACGAAGGTGATGGATGGACACACTGTACTAGGC 60
|||||
Db 25 CTGCTGCTCGAGGAGCGTGGAGCCGACGAAGGTGATGGATGGACACACTGTACTAGGC 84

RESULT 2

H85099
LOCUS Ys70e05.sl Soares retina N2b4HR Homo sapiens CDNA clone
DEFINITION IMAGE:220160 3', mRNA sequence.
ACCESSION H85099.1 GI:1063842
VERSION EST.
KEYWORDS human.
SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

TITLE

JOURNAL
COMMENT

High quality sequence stops: 297

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 860 Std Error: 0.00

Seq primer: Promega -21m13

High quality sequence stop: 297.

FEATURES

source

Location/Qualifiers
1. .388
/organism="Homo sapiens"
/db_xref="GDB:3848153"
/db_xref="taxon:9606"
/clone="IMAGE:220160"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"

/note="Organ: eye; Vector: pT73D (Pharmacia) with a modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Patima Bonaldo."

BASE COUNT 91 a 93 c 100 g 102 t 2 others

ORIGIN

Query Match 100.0%; Score 60; DB 14; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.9e-21;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTGCTCGAGGAGCGTGGAGCCGACGAAGGTGATGGATGGACACACTGTACTAGGC 60
|||||
Db 71 CTGCTGCTCGAGGAGCGTGGAGCCGACGAAGGTGATGGATGGACACACTGTACTAGGC 130

RESULT 3

H99544
LOCUS Yx29g09.sl Soares melanocyte 2NBHM Homo sapiens CDNA clone
DEFINITION IMAGE:263200 3', mRNA sequence.
ACCESSION H99544
VERSION H99544.1 GI:1124212
KEYWORDS EST.
SOURCE human.
ORGANISM

REFERENCE

AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

TITLE

JOURNAL
COMMENT

High quality sequence stops: 296

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: polyT not found

Insert Length: 1174 Std Error: 0.00

Seq primer: m13 -40 forward

High quality sequence stop: 296.

Location/Qualifiers

1. .418

/organism="Homo sapiens"

/db_xref="GDB:3872842"

```

/db_xref="taxon:9606"
/clone="IMAGE:263200"
/clone_lib="Soares melanocyte 2NBHM"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="PH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCGAAGTGGGAGCGGCCGAGTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino."
          91 a    117 c   97 g   108 t      5 others
BASE COUNT
ORIGIN

Query Match       100.0%; Score 60; DB 14; Length 418;
Best Local Similarity 100.0%; Pred. No. 3e-21;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTGCTCAGGACGCTGGAGCCGCACGAGGTGATGGATGTGGACACACTGTACTAGGC 60
|||||
Db 115 CTGCTGCTCAGGACGCTGGAGCCGCACGAGGTGATGGATGTGGACACACTGTACTAGGC 174
|||||

RESULT 4
AA885289
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA885289
al58q05.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1461560 3' similar to TR:Q14089 Q14089 HYPOTHETICAL PROTEIN
: mRNA sequence.
AA885289
AA885289.1 GI:2994366
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 459)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand
Insert length: 1052 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 450.
Location/Qualifiers
1..459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1461560"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19w, testis NHR, and B-cell
NCI_CCAP_GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
```



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:209450"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/lab_host="DH10B"
/dev_stage="adult"
/organ="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      130 a   130 c   111 g   154 t      1 others
ORIGIN

Query Match      100.0%; Score 60; DB 9; Length 526;
Best Local Similarity 100.0%; Pred. No. 3.3e-21;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTCTCGAGGACGCTGAGCCAGCGATGGATGGATGGACACACTGTACTAGGC 60
|||||
Db 97 CTGCTCTCGAGGACGCTGAGCCAGCGATGGATGGATGGACACACTGTACTAGGC 156
|||||

RESULT 9
BF434973
LOCUS
DEFINITION
7p04a07.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644820 3'
similar to TR:Q14089 Q14089 HYPOTHETICAL 40.0 KD PROTEIN ;, mRNA
sequence.
BF434973
BF434973.1 GI:11447261
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 566)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Cloned distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 501.
Location/Qualifiers
1. 566
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3644820"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrothecoma"
/lab_host="PH10B (phage-resistant)"
/organ="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTCAAGTGGAGCGCGCGACATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      138 a   146 c   120 g   162 t
ORIGIN

Query Match      100.0%; Score 60; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 3.4e-21;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTCTCGAGGACGCTGAGCCAGCGATGGATGGATGGACACACTGTACTAGGC 60
|||||
Db 106 CTGCTCTCGAGGACGCTGAGCCAGCGATGGATGGATGGACACACTGTACTAGGC 165
|||||

RESULT 10
AW138866
LOCUS
DEFINITION
UI-H-B11-aeq-b-04-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
IMAGE:2720190 3', mRNA sequence.
AW138866
AW138866
AW138866.1 GI:6143184
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 605)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Oligo-dT track not found. Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
www-bio.llnl.gov/bbrp/image/image.html The following repetitive
elements were found in this cDNA sequence: 1-26,
>GC rich#Low-complexity 27-122, >(GGA)n#Simple_repeat
Seq primer: M13 Forward
POLYA-No. Location/Qualifiers
1. 605
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2720190"
/clone_lib="NCI_CGAP_Sub3"
/lab_host="DH10B (Life Technologies)"
/organ="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; The
NCI_CGAP_Sub3 library is a subtracted library derived from
the NCI_CGAP_Sub1 library, which is a subtracted library
derived from BI. BI constitutes a mixture of 21
normalized or subtracted NCI_CGAP libraries: NCI_CGAP_Co4
, NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10,
NCI_CGAP_Co16, NCI_CGAP_Kid5, NCI_CGAP_Kid12,
NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2,
NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1, NCI_CGAP_Le12,
NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24,
NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
NCI_CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clonids 1322376-1323911, 1456008-1456775
, 1500552-1502855); NCI_CGAP_Kid5 pool 1 LLAM 3338-3342
, 3722-3725, 3776-3778 (IMAGE Clonids 1323912-1325831,
1471368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1
LLAM 3575-3582, 3851-3854 (IMAGE Clonids 1414920-1417991,
1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clonids 1257096-1258631,

```

1469064-1470983, 1475592-1476743}; NCI_CGAP_Pr22 pool 1
 LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones)
 985608-986759, 1101192-1101959, 1217928-1220615};
 NCI_CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE
 Clones) 1057416-1061255, 1144584-1145351}. Subtraction
 was performed as previously described [Bonaldi, Lennon &
 Soares (1996)]: Normalization and Subtraction: Two
 Approaches To Facilitate Gene Discovery. Genome Research
 6, 791-806.
 TAG_LIB=NCI_CGAP_kid3
 TAG_ISSUE=Kidney
 TAG_SEQ=AATGC"

BASE COUNT 142 a 157 c 132 g 174 t
 ORIGIN

Query Match 100.0%; Score 60; DB 10; Length 605;
 Best Local Similarity 100.0%; Pred. No. 3.5e-21;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTGCTCGAGGAGCTGGAGCCAGCAAGGTGATGGTGGACACACTGTACTAGGC 60
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 114 CTGCTGCTCGAGGAGCTGGAGCCAGCAAGGTGATGGTGGACACACTGTACTAGGC 173

RESULT 11
 BF732384
 LOCUS
 DEFINITION naellia2.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3434806 3',
 similar to TR:Q14089 Q14089 HYPOTHETICAL 40.0 KD PROTEIN ;, mRNA
 sequence.
 ACCESSION BF732384
 VERSION BF732384.1 GI:12057542
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 693)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
 Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov

Seq primer: -400P from Gibco
 High quality sequence stop: 490.

FEATURES
 Location/Qualifiers
 1..693
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3434806"
 /clone_lib="NCI_CGAP_Ov18"
 /tissue_type="fibrotheoma"
 /lab_host="DH10B (phage-resistant)"

/note="organ: ovary; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTACCACTGTAAGTGGAGCGCCGCCGACATTTTTTTTTTTT 3'];
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldi."

BASE COUNT 162 a 173 c 159 g 197 t 2 others
 ORIGIN

Query Match 100.0%; Score 60; DB 12; Length 693;
 Best Local Similarity 100.0%; Pred. No. 3.6e-21;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTGCTCGAGGAGCTGGAGCCAGCAAGGTGATGGTGGACACACTGTACTAGGC 60
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 115 CTGTGCTCGAGGAGCTGGAGCCAGCAAGGTGATGGTGGACACACTGTACTAGGC 174

RESULT 12
 AW070902
 LOCUS

DEFINITION xa31b09.x1 NCI_CGAP_Br18 Homo sapiens cDNA clone IMAGE:2568377 3',
 similar to TR:Q14089 Q14089 HYPOTHETICAL 40.0 KD PROTEIN ;, mRNA
 sequence.

ACCESSION AW070902
 VERSION AW070902.1 GI:6025900
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 703)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life

Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand

Insert Length: 1439 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 436.

FEATURES
 Location/Qualifiers

1..703
 /organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2568377"

/clone_lib="NCI_CGAP_Br18"

/tissue_type="four pooled high-grade tumors, including two

primary tumors and two metastatic to ovary"

/lab_host="DH10B"

/note="organ: breast; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies."

BASE COUNT 165 a 176 c 158 g 204 t

ORIGIN

Query Match 100.0%; Score 60; DB 10; Length 703;
 Best Local Similarity 100.0%; Pred. No. 3.7e-21;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGTGCTCGAGGAGCTGGAGCCAGCAAGGTGATGGTGGACACACTGTACTAGGC 60
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 107 CTGTGCTCGAGGAGCTGGAGCCAGCAAGGTGATGGTGGACACACTGTACTAGGC 166

RESULT 13

BG178371

LOCUS

DEFINITION 602330141F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4431275 5',
 mRNA sequence.

ACCESSION BG178371

```

VERSION BGI78371.1 GI:12685074
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
cDNA Library Preparation: Life Technologies, Inc.
Tissue Procurement: DCTD/DPD
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Invitrogen, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10186 row: b column: 12
High quality sequence stop: 2
High quality sequence start: 654.
High quality sequence stop: 654.
Location/Qualifiers
1..908
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 229 a 219 c 216 g 242 t 2 others
ORIGIN
Query Match 100.0%; Score 60; DB 12; Length 908;
Best Local Similarity 100.0%; Pred. No. 4.1e-21;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTGCTCGAGGACGCTGGAGCCAGCAAGGTGATGGATGGACACTGTACTAGGC 60
|||||
Db 57 CTGCTGCTCGAGGACGCTGGAGCCAGCAAGGTGATGGATGGACACTGTACTAGGC 116
|||||

RESULT 14
BM905706
LOCUS BM905706
DEFINITION AGENCOURT_6721242 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5556057
5', mRNA sequence.
ACCESSION BM905706
VERSION BM905706.1 GI:19355085
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12277 row: d column: 10

```

```

FEATURES
source
High quality sequence stop: 629.
Location/Qualifiers
1..1061
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb.
BASE COUNT 281 a 241 c 255 g 283 t 1 others
ORIGIN
Query Match 100.0%; Score 60; DB 14; Length 1061;
Best Local Similarity 100.0%; Pred. No. 4.3e-21;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTGCTCGAGGACGCTGGAGCCAGCAAGGTGATGGATGGACACTGTACTAGGC 60
|||||
Db 76 CTGCTGCTCGAGGACGCTGGAGCCAGCAAGGTGATGGATGGACACTGTACTAGGC 135
|||||

RESULT 15
BF732861
LOCUS BF732861
DEFINITION nael6d03.xl NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:343517 3',
sequence.
ACCESSION BF732861
VERSION BF732861.1 GI:12057936
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor gene index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento
Soares, Ph.D. CDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 459.
Location/Qualifiers
1..692
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrotheoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAGTGGAGCGCGCGGACATTTTGTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 164 a 177 c 153 g 198 t
ORIGIN
FEATURES
source
High quality sequence stop: 459.
Location/Qualifiers
1..692
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrotheoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAGTGGAGCGCGCGGACATTTTGTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 164 a 177 c 153 g 198 t
ORIGIN

```

Query Match 76.7%; Score 46; DB 12; Length 592;
Best Local Similarity 100.0%; Pred. NO. 8.4e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 15 CGCTGGAGCCCGCAGCAAGGTGATGGATGTGGACACACTGTACTAGGC 60
|||||
Db 128 CGCTGGAGCCCGCAGCAAGGTGATGGATGTGGACACACTGTACTAGGC 173
|||||

Search completed: January 21, 2003, 14:43:27
Job time : 304.512 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 13:00:06 ; Search time 37.2441 Seconds
(without alignments)
3292.167 Million cell updates/sec

Title: US-10-060-830-6
Perfect score: 275
Sequence: 1 caacttcagtgtgtagcc.....tttgaagatgatgtgttt 275

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 393868 seqs, 222934149 residues

Word size : 0 787736

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PublishedApplications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	580	10 US-09-864-761-12400	Sequence 12400, A
2	275	100.0	5657	9 US-09-974-298-96	Sequence 96, Appl
3	257	93.5	467	10 US-09-864-761-30344	Sequence 30344, A
4	19	6.9	9417	10 US-09-764-855-171	Sequence 171, App
5	18	6.5	9377	10 US-09-801-874-3	Sequence 3, Appli
6	17	6.2	456	10 US-09-924-0358-677	Sequence 677, App
7	17	6.2	460	10 US-09-880-107-537	Sequence 537, App
8	17	6.2	577	10 US-09-864-761-12564	Sequence 12564, A
9	17	6.2	613	10 US-09-770-149-855	Sequence 855, App
10	17	6.2	1230	10 US-09-764-847-1854	Sequence 1854, Ap
11	17	6.2	2529	10 US-09-764-847-1859	Sequence 1859, Ap
12	17	6.2	3126	10 US-09-815-242-4020	Sequence 4020, Ap
13	17	6.2	4138	10 US-09-802-472B-5	Sequence 5, Appli
14	17	6.2	5814	10 US-09-764-847-1860	Sequence 1860, Ap
15	17	6.2	26048	10 US-09-764-869-1556	Sequence 1556, Ap
16	17	6.2	26591	10 US-09-764-877-2678	Sequence 2678, Ap
17	16	5.8	189	10 US-09-864-761-31208	Sequence 31208, A
18	16	5.8	213	10 US-09-974-300-7540	Sequence 7540, Ap
19	16	5.8	223	9 US-09-954-531-677	Sequence 677, App

20	16	5.8	223	9	US-09-954-531-1092	Sequence 1092, Ap
21	16	5.8	223	10	US-09-969-708-177	Sequence 177, App
c 22	16	5.8	279	10	US-09-960-352-11414	Sequence 11414, A
23	16	5.8	356	10	US-09-604-287A-476	Sequence 476, App
24	16	5.8	356	12	US-10-007-805-476	Sequence 476, App
c 25	16	5.8	377	9	US-09-736-457-1096	Sequence 1096, Ap
c 26	16	5.8	377	9	US-09-902-941-1096	Sequence 1096, Ap
c 27	16	5.8	377	9	US-09-849-626-1096	Sequence 1096, Ap
28	16	5.8	456	10	US-09-864-761-16100	Sequence 16100, A
29	16	5.8	477	10	US-09-864-761-14675	Sequence 14675, A
c 30	16	5.8	479	10	US-09-864-761-3875	Sequence 3875, Ap
c 31	16	5.8	577	10	US-09-864-761-12010	Sequence 12010, A
c 32	16	5.8	634	10	US-09-764-855-15	Sequence 15, Appl
c 33	16	5.8	659	10	US-09-879-536-656	Sequence 656, App
34	16	5.8	717	10	US-09-925-300-1	Sequence 1, Appli
35	16	5.8	725	10	US-09-770-149-129	Sequence 129, App
c 36	16	5.8	823	10	US-09-925-299-145	Sequence 145, App
37	16	5.8	1190	9	US-09-962-832-96	Sequence 96, Appl
38	16	5.8	1299	9	US-10-001-987-31	Sequence 31, Appl
39	16	5.8	1366	10	US-09-864-761-32593	Sequence 32593, A
40	16	5.8	1382	10	US-09-925-300-692	Sequence 692, App
c 41	16	5.8	1656	10	US-09-925-301-125	Sequence 125, App
42	16	5.8	1817	9	US-09-764-868-68	Sequence 68, Appl
43	16	5.8	1876	10	US-09-604-287A-477	Sequence 477, App
44	16	5.8	1876	12	US-10-007-805-477	Sequence 477, App
45	16	5.8	2396	10	US-09-748-107-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-864-761-12400
; Sequence 12400, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Genp, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30

102e

```
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12400
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
; US-09-864-761-12400

Query Match      100.0%; Score 275; DB 10; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.7e-134;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC 60
Db 281 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC 340
QY 61 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCCAGCCAGG 120
Db 341 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCCAGCCAGG 400
QY 121 CCCAGTATGATACCCCGAAGCTGGGAAGCCAGGTCTACCTGCCCGACAGCAATGGTGT 180
Db 401 CCCAGTATGATACCCCGAAGCTGGGAAGCCAGGTCTACCTGCCCGACAGCAATGGTGT 460
QY 181 ACCAGTGCCACAGACACACAAGATATCAGGACGAGGAGGATGGGGAATGTGATG 240
Db 461 ACCAGTGCCACAGACACACAAGATATCAGGACGAGGAGGATGGGGAATGTGATG 520
QY 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275
Db 521 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 555

RESULT 2
US-09-974-298-96
; Sequence 96, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/236,331
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 96
; LENGTH: 5657
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 241227.17
; NAME/KEY: unsure
; LOCATION: 4516
; OTHER INFORMATION: a, t, c, g, or other
; US-09-974-298-96
```

```
Query Match      100.0%; Score 275; DB 9; Length 5657;
Best Local Similarity 100.0%; Pred. No. 3.1e-134;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC 60
Db 1931 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCC 1990
QY 61 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCCAGCCAGG 120
Db 1991 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCCAGCCAGG 2050
QY 121 CCCAGTATGATACCCCGAAGCTGGGAAGCCAGGTCTACCTGCCCGACAGCAATGGTGT 180
Db 2051 CCCAGTATGATACCCCGAAGCTGGGAAGCCAGGTCTACCTGCCCGACAGCAATGGTGT 2110
QY 181 ACCAGTGCCACAGACACACAAGATATCAGGACGAGGAGGATGGGGAATGTGATG 240
Db 2111 ACCAGTGCCACAGACACACAAGATATCAGGACGAGGAGGATGGGGAATGTGATG 2170
QY 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275
Db 2171 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 2205

RESULT 3
US-09-864-761-30344
; Sequence 30344, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aescmca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00659
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
```

;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 30344
;; LENGTH: 467
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC013497.4
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
;; OTHER INFORMATION: EST_HUMAN HIT: AW583777.1, EVALUE 3.00e-51
;; OTHER INFORMATION: NT HIT: Z70177.1, EVALUE 2.90e-01
;; OTHER INFORMATION: SWISSPROT HIT: Q9Y011, EVALUE 7.40e-01
US-09-864-761-30344

Query Match 93.5%; Score 257; DB 10; Length 467;
Best Local Similarity 100.0%; Pred. No. 6.9e-125;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACTTCAGTGGTGGTCCGAGCCCTCCACATCCACTTTCAGGCTACGGGGAACCAACCTCCCC 60
|||||
DB 211 CCACTTCAGTGGTGGTCCGAGCCCTCCACATCCACTTTCAGGCTACGGGGAACCAACCTCCCC 270
|||||
QY 61 CACTAGTGGGAACCTTACAAATACACTTCTCTCCAGACTGACAGCTGCTCTCCAGCCAGG 120
|||||
DB 271 CACTAGTGGGAACCTTACAAATACACTTCTCTCCAGACTGACAGCTGCTCTCCAGCCAGG 330
|||||
QY 121 CCCAGTATGATACCCGAAAGCTGGGAAGCCAGGCTTACCTCCCCAGACGAAATTGGTGT 180
|||||
DB 331 CCCAGTATGATACCCGAAAGCTGGGAAGCCAGGCTTACCTCCCCAGACGAAATTGGTGT 390
|||||
QY 181 ACCAGTCCAGAGCAGACACAGAGATGATCAGGAGCAGGAGGATGGGAATGTGTATG 240
|||||
DB 391 ACCAGTCCAGAGCAGACACAGAGATGATCAGGAGCAGGAGGATGGGAATGTGTATG 450
|||||
QY 241 TTTTAAAGAAATCCTT 257
|||||
DB 451 TTTTAAAGAAATCCTT 467
|||||

RESULT 4
US-09-764-855-171
;; Sequence 171, Application US/09764855
;; Patent No. US20020119919A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: P110
;; CURRENT APPLICATION NUMBER: US/09/764,855
;; CURRENT FILING DATE: 2001-01-17
;; Prior application data removed - consult PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 334
;; SOFTWARE: PatentIn ver. 2.0
;; SEQ ID NO 171
;; LENGTH: 9417
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (3)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: SITE
;; LOCATION: (8736)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: SITE
;; LOCATION: (8737)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: SITE
;; LOCATION: (8738)

;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: SITE
;; LOCATION: (8765)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: SITE
;; LOCATION: (8769)
;; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-855-171

Query Match 6.9%; Score 19; DB 10; Length 9417;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TCAGTTGGTCAGCCCTCCCA 24
|||||
DB 4701 TCAGTTGGTCAGCCCTCCCA 4719
|||||

RESULT 5
US-09-801-874-3
;; Sequence 3, Application US/09801874
;; Patent No. US2002004801A1
;; GENERAL INFORMATION:
;; APPLICANT: YAN Chunhua et al.
;; TITLE OF INVENTION: ISOLATED HUMAN AMINOTRANSFERASE
;; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
;; TITLE OF INVENTION: AMINOTRANSFERASE PROTEINS,
;; TITLE OF INVENTION: AND USES THEREOF
;; FILE REFERENCE: C1000615
;; CURRENT APPLICATION NUMBER: US/09/801,874
;; CURRENT FILING DATE: 8001-03-09
;; NUMBER OF SEQ ID NOS: 5
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 9377
;; TYPE: DNA
;; ORGANISM: Human
US-09-801-874-3

Query Match 6.5%; Score 18; DB 10; Length 9377;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 213 GGAGCAGGAAGGATGGG 230
|||||
DB 6424 GGAGCAGGAAGGATGGG 6441
|||||

RESULT 6
US-09-924-035A-677/c
;; Sequence 677, Application US/09924035A
;; Patent No. US20020142319A1
;; GENERAL INFORMATION:
;; APPLICANT: Griach, Jrn
;; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
;; TITLE OF INVENTION: thaliana
;; FILE REFERENCE: 2011US
;; CURRENT APPLICATION NUMBER: US/09/924,035A
;; CURRENT FILING DATE: 2000-08-11
;; PRIOR APPLICATION NUMBER: US 60/148,784
;; PRIOR FILING DATE: 1999-08-13
;; NUMBER OF SEQ ID NOS: 900
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 677
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(456)
;; OTHER INFORMATION: n = A,T,C or G
US-09-924-035A-677

Query Match 6.2%; Score 17; DB 10; Length 456;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 226 ATGGGAATGTGATGTT 242
 |||||
 Db 94 ATGGGAATGTGATGTT 78

RESULT 7

US-09-880-107-537/c
 ; Sequence 537, Application US/09880107
 ; Patent No. US20020142981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Horne, Darci T.
 ; APPLICANT: Vockley, Joseph G.
 ; APPLICANT: Scherf, Uwe
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 ; FILE REFERENCE: 44921-5028-WO
 ; CURRENT APPLICATION NUMBER: US/09/880,107
 ; CURRENT FILING DATE: 2001-05-14
 ; PRIOR APPLICATION NUMBER: US 60/211,379
 ; PRIOR FILING DATE: 2000-06-14
 ; PRIOR APPLICATION NUMBER: US 60/237,054
 ; PRIOR FILING DATE: 2000-10-02
 ; NUMBER OF SEQ ID NOS: 3950
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 537
 ; LENGTH: 460
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA250775
 US-09-880-107-537

Query Match 6.2%; Score 17; DB 10; Length 460;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 236 TGATGTTTTTAAGAAA 252
 |||||
 Db 127 TGATGTTTTTAAGAAA 111

RESULT 8

US-09-864-761-12564
 ; Sequence 12564, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aemica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 12564
 ; LENGTH: 577
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC010492.4
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.77
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.79
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.81
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
 US-09-864-761-12564

Query Match 6.2%; Score 17; DB 10; Length 577;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CTCTCTCCAGGACTGA 100
 |||||
 Db 45 CTCTCTCCAGGACTGA 61

RESULT 9

US-09-770-149-855
 ; Sequence 855, Application US/09770149
 ; Patent No. US20020059663A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gorlach, Jorn
 ; APPLICANT: An, Yong-Qiang
 ; APPLICANT: Hamilton, Carol M.
 ; APPLICANT: Price, Jennifer L.
 ; APPLICANT: Raines, Tracy M.
 ; APPLICANT: Yu, Yang
 ; APPLICANT: Rameaka, Joshua G.
 ; APPLICANT: Page, Amy
 ; APPLICANT: Matthew, Abraham V.
 ; APPLICANT: Ledford, Brooke L.
 ; APPLICANT: Woessner, Jeffrey P.
 ; APPLICANT: Haas, William David
 ; APPLICANT: Garcia, Carlos A.
 ; APPLICANT: Kricker, Maja
 ; APPLICANT: Slader, Ted
 ; APPLICANT: Davis, Keith R.
 ; APPLICANT: Allen, Keith
 ; APPLICANT: Hoffman, Neil
 ; APPLICANT: Hurban, Patrick
 ; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
 ; TITLE OF INVENTION: thaliana

FILE REFERENCE: 2024 (PARA-013PRV)
CURRENT APPLICATION NUMBER: US/09/770,149
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,506
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 855
LENGTH: 613
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-149-855

Query Match 6.2%; Score 17; DB 10; Length 613;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 226 ATGGGGAATGTGATGTT 242
|||||
Db 512 ATGGGGAATGTGATGTT 528

RESULT 10

US-09-764-847-1854/c
Sequence 1854, Application US/09764847
Patent No. US20020132767A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009
CURRENT APPLICATION NUMBER: US/09/764,847
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 2003
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1854
LENGTH: 1230
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-847-1854

Query Match 6.2%; Score 17; DB 10; Length 1230;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 226 ATGGGGAATGTGATGTT 242
|||||
Db 491 ATGGGGAATGTGATGTT 475

RESULT 11

US-09-764-847-1859/c
Sequence 1859, Application US/09764847
Patent No. US20020132767A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009
CURRENT APPLICATION NUMBER: US/09/764,847
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 2003
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1859
LENGTH: 2529
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-847-1859

Query Match 6.2%; Score 17; DB 10; Length 2529;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 226 ATGGGGAATGTGATGTT 242
|||||
Db 1790 ATGGGGAATGTGATGTT 1774

RESULT 12

US-09-815-242-4020
Sequence 4020, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4020
LENGTH: 3126
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4020

Query Match 6.2%; Score 17; DB 10; Length 3126;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 AAGCCAGGTCTACCTGC 163
|||||
Db 2022 AAGCCAGGTCTACCTGC 2038

RESULT 13

US-09-802-472B-5
Sequence 5, Application US/09802472B
Patent No. US20020103353A1
GENERAL INFORMATION:
APPLICANT: EINAT, Paz
APPLICANT: SKALITER, Rami
APPLICANT: FEINSTEIN, Elena
TITLE OF INVENTION: SEQUENCES CHARACTERISTIC OF HYPOXIA-REGULATED GENE TRANSCRIPT
FILE REFERENCE: EINAT-4.1C
CURRENT APPLICATION NUMBER: US/09/802,472B
CURRENT FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: US 09/383,096
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: US 09/138,109
PRIOR FILING DATE: 1998-08-21
PRIOR APPLICATION NUMBER: US 60/098,158
PRIOR FILING DATE: 1998-08-27
PRIOR APPLICATION NUMBER: US 60/132,684

; PRIOR FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 4138
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-802-472B-5

Query Match 6.2%; Score 17; DB 10; Length 4138;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 TGATGTTTTTAAAGAAA 252
|||||
Db 3111 TGATGTTTTTAAAGAAA 3127

RESULT 14

US-09-764-847-1860/c
; Sequence 1860, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1860
; LENGTH: 5814
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1860

Query Match 6.2%; Score 17; DB 10; Length 5814;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 ATGGGGAATGTGATGTT 242
|||||
Db 5075 ATGGGGAATGTGATGTT 5059

RESULT 15

US-09-764-869-1556/c
; Sequence 1556, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1556
; LENGTH: 26048
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-1556

Query Match 6.2%; Score 17; DB 10; Length 26048;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 CCTAGCCCGGCCGAG 125
|||||
Db 9740 CCTAGCCCGGCCGAG 9724

Search completed: January 21, 2003, 14:46:45
Job time : 46.2441 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 10:13:26 ; Search time 175.276 Seconds
(without alignments)
3854.498 Million cell updates/sec

Title: us-10-060-830-4
Perfect score: 300
Sequence: 1 aaacagaaagaaaaaactg.....tatgcagacctagatcctta 300

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	300	100.0	2046	21 AAZ51872	Human Factor 8 Hom
2	230	76.7	594	23 AAS65485	DNA encoding novel
3	183	61.0	640	21 AAZ80396	Human colon cancer
4	109	36.3	580	22 ABA60693	Human foetal liver
5	109	36.3	580	22 AAK08975	Human brain expres
6	109	36.3	580	22 AAK34866	Human bone marrow
7	109	36.3	580	22 AAI40582	Probe #9268 used t
8	106	35.3	467	22 ABA76819	Human foetal liver
9	106	35.3	467	22 AAK25439	Human brain expres

10	106	35.3	467	22 AAK51451	Human bone marrow
11	106	35.3	467	22 AA157541	Probe #26227 used
12	58	19.3	1907	24 AAD35994	Human neuropilin-H
13	58	19.3	1907	24 ABK49567	Human CDNA encoding
14	58	19.3	1907	24 ABN59799	Novel human coding
15	58	19.3	2428	22 AAH34855	Human colon cancer
16	20	6.7	393	24 ABN66437	Streptococcus poly
17	20	6.7	440	21 AAF11554	Aspergillus niger
18	20	6.7	11091	24 ABN92780	Staphylococcus epi
19	19	6.3	423	22 ABA48060	Human breast cell
20	19	6.3	423	22 ABA65943	Human foetal liver
21	19	6.3	423	22 ABA33035	Probe #11501 for g
22	19	6.3	423	22 AAK14364	Human brain expres
23	19	6.3	423	22 AAK40093	Human bone marrow
24	19	6.3	423	22 AAI20873	Probe #10806 for g
25	19	6.3	423	22 AAI46111	Probe #14797 used
26	19	6.3	423	22 AAI06583	Probe #6574 used t
27	19	6.3	423	24 ABS14168	Human genome-deriv
28	19	6.3	1497	22 AAK87499	Human immune/haema
29	19	6.3	2122	23 AAS80960	DNA encoding novel
30	19	6.3	2521	23 AAS66088	DNA encoding novel
31	19	6.3	7105	22 AAL37450	Human musculoskele
32	19	6.3	13585	17 AAT11549	Tumour rejection a
33	19	6.3	30365	22 AAD20405	P. syringae pv. to
34	18	6.0	286	21 AAC53732	Arabidopsis thalia
35	18	6.0	370	24 ABL77512	Human ovarian canc
36	18	6.0	452	24 ABL93879	Arabidopsis thalia
37	18	6.0	581	22 AAI44194	Probe #12880 used
38	18	6.0	789	20 AAX13443	Enterococcus faeca
39	18	6.0	2708	22 AAI7390	Human CDNA sequenc
40	18	6.0	3232	21 ABL58202	Polyhydroxyalkanoa
41	18	6.0	4240	24 ABI99878	Mouse ischaemic co
42	18	6.0	4812	22 AAL04812	Human reproductive
43	18	6.0	4812	23 ABL97706	Human testicular a
44	18	6.0	13021	22 AAL04666	Human reproductive
45	18	6.0	13021	23 ABL97573	Human testicular a

ALIGNMENTS

RESULT 1
ID AAZ51872 standard; cDNA; 2046 BP.
XX AAZ51872;
XX AC AAZ51872;
XX DT 04-JUL-2000 (first entry)
XX DE Human Factor 8 Homologue cDNA.
XX KW Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
XX KW cerebroprotective; therapeutic; coagulation related disorder;
XX KW haemophilia; stroke; screening; ss.
XX OS Homo sapiens.
XX FH Key
XX CDS Location/Qualifiers
FT 19..2028
FT /*tag= a
FT /product= "Factor 8 Homologue"

PN WO200012532-A1.

XX 09-MAR-2000.

XX PF 20-AUG-1999; 99WO-US19047.

XX PR 31-AUG-1998; 98US-0098521.

XX PA (ELIL) LILLY & CO ELI.

XX PR Rosteck PRJ, Su W, Li XM;


```
KW hyperplasia; ds.
XX
OS Homo sapiens.
XX
PN WO9964576-A2.
XX
XX 16-DEC-1999.
XX
XX 09-JUN-1999; 99WO-IB01062.
XX
XX 10-JUN-1998; 98US-0088801.
XX
XX (FARB ) BAYER CORP.
XX
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX
XX WPI; 2000-087220/07.
XX
XX Novel nucleic acids, used to develop products for the diagnosis and
XX treatment of disorders involving unwanted cell proliferation,
XX particularly cancers, especially colon cancer.
XX
XX Claim 15; Page 322; 469pp; English.
XX
XX AA79917 to AA280766 represent double stranded cDNA clones isolated from
XX the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
XX cDNA clones can be used to generate antisense oligonucleotides which
XX can be used for antisense therapy. Methods and products from the present
XX invention can be used for identifying and/or classifying cancerous cells
XX present in a human tumour, particularly in solid tumours, e.g.
XX carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
XX can be used for developing agents for the diagnosis and treatment of
XX disorders involving unwanted cell proliferation, such as neoplasia,
XX dysplasia or hyperplasia.
XX
XX Sequence 640 BP; 148 A; 134 C; 160 G; 192 T; 6 other;
XX
XX Query Match 61.0%; Score 183; DB 21; Length 640;
XX Best Local Similarity 100.0%; Pred. No. 2.9e-79;
XX Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 49 CGGGCAGTTGGTGGAAAGGAATGAAGCAGTTCTTCTGCAAAAGCAGTGGACCATGA 108
XX |||||
XX Db 184 CGGGCAGTTGGTGGAAAGGAATGAAGCAGTTCTTCTGCAAAAGCAGTGGACCATGA 125
XX |||||
XX QY 109 GGAACCCAGTTCCCTATAGCAGCAGGAAGTTTATCACCTGAGTCCAGAGAGAGTCAC 168
XX |||||
XX Db 124 GGAACCCAGTTCCCTATAGCAGCAGGAAGTTTATCACCTGAGTCCAGAGAGAGTCAC 65
XX |||||
XX QY 169 CACAGTGTGCGAGGTGACTCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATTGTTGG 228
XX |||||
XX Db 64 CACAGTGTGCGAGGTGACTCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATTGTTGG 5
XX |||||
XX QY 229 TAC 231
XX |||
XX Db 4 TAC 2
XX
XX RESULT 4
XX ABA60693
XX ID ABA60693 standard; DNA; 580 BP.
XX
XX AC ABA60693;
XX
XX XX 01-FEB-2002 (first entry)
XX
XX DE Human foetal liver single exon nucleic acid probe #8998.
XX
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX OS Homo sapiens.
```

```
XX WO200157277-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00669.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver.
XX
XX PS Claim 1; SEQ ID NO 8998; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;
XX
XX Query Match 36.3%; Score 109; DB 22; Length 580;
XX Best Local Similarity 100.0%; Pred. No. 3.1e-43;
XX Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 192 CAGAGTATGCTCAGCCACTGGTAGGAGGAATGTTGGTACACTTCATCAAGATCTACCT 251
XX |||||
XX Db 68 CAGAGTATGCTCAGCCACTGGTAGGAGGAATGTTGGTACACTTCATCAAGATCTACCT 127
XX |||||
XX QY 252 TTAACCCAGCAAGAAAGAAAGCAGGCTATGCAGACTAGATCCTTA 300
XX |||||
XX Db 128 TTAACCCAGCAAGAAAGAAAGCAGGCTATGCAGACTAGATCCTTA 176
XX |||||
XX
XX RESULT 5
XX AAK08975
XX ID AAK08975 standard; DNA; 580 BP.
XX
XX AC AAK08975;
XX
XX XX 05-NOV-2001 (first entry)
XX
XX DE Human brain expressed single exon probe SEQ ID NO: 8966.
XX
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157275-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00667.
XX
XX
```


CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;

SQ Query Match 36.3%; Score 109; DB 22; Length 580;

Best Local Similarity 100.0%; Pred. No. 3.1e-43;

Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 CAGAGTATGCTCAGCCACTGCTAGAGGAATTGTTGGTACACTTCATCAAGATCTACCT 251

Db 68 CAGAGTATGCTCAGCCACTGCTAGAGGAATTGTTGGTACACTTCATCAAGATCTACCT 127

QY 252 TTAACACAGAGAGGAAAGAAAGCAGGCTATGCAGACTAGATCCTTA 300

Db 128 TTAACACAGAGAGGAAAGAAAGCAGGCTATGCAGACTAGATCCTTA 176

RESULT 8

ABA76819

ID ABA76819 standard; DNA; 467 BP.

XX ABA76819;

AC ABA76819;

XX 01-FEB-2002 (first entry)

DT Human foetal liver single exon nucleic acid probe #25124.

DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

OS WO200157277-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

XX Claim 4; SEQ ID NO 25124; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX foetal liver. The present sequence is a single exon nucleic acid

XX probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;

XX Query Match 35.3%; Score 106; DB 22; Length 467;

XX Best Local Similarity 100.0%; Pred. No. 9.1e-42;

XX Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 AGTATGCTCAGCCACTGCTAGAGGAATTGTTGGTACACTTCATCAAGATCTACCTTTA 254

Db 1 AGTATGCTCAGCCACTGCTAGAGGAATTGTTGGTACACTTCATCAAGATCTACCTTTA 60

QY 255 AACACAGAGAGGAAAGAAAGCAGGCTATGCAGACTAGATCCTTA 300

Db 61 AACACAGAGAGGAAAGAAAGCAGGCTATGCAGACTAGATCCTTA 106

RESULT 9

AAK25439

ID AAK25439 standard; DNA; 467 BP.

XX AAK25439;

XX 05-NOV-2001 (first entry)

DT Human brain expressed single exon probe SEQ ID NO: 25430.

XX Human; brain expressed exon; gene expression analysis; probe;

XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX epilepsy; cancer; ss.

XX Homo sapiens.

OS WO200157277-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX brains -

XX Example 4; SEQ ID NO: 25430; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is one of the probes of the

XX invention.

XX SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;

XX Query Match 35.3%; Score 106; DB 22; Length 467;

XX Best Local Similarity 100.0%; Pred. No. 9.1e-42;

XX Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 AGTATGCTCAGCCACTGCTAGAGGAATTGTTGGTACACTTCATCAAGATCTACCTTTA 254

Db 1 AGTATGCTCAGCCACTGCTAGAGGAATTGTTGGTACACTTCATCAAGATCTACCTTTA 60

QY 255 AACACAGAGAGGAAAGAAAGCAGGCTATGCAGACTAGATCCTTA 300

Db 61 AACACAGAGAGGAAAGAAAGCAGGCTATGCAGACTAGATCCTTA 106

```

RESULT 10
AAK51451
ID AAK51451 standard; DNA; 467 BP.
XX
AC AAK51451;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 26008.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 26008; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;

Query Match 35.3%; Score 106; DB 22; Length 467;
Best Local Similarity 100.0%; Pred. No. 9.1e-42;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 AGTATGCTCAGCCACTGGTAGGAGAAATGTTGGTACATCATCAAGATCTACCTTTA 254
DB 1 AGTATGCTCAGCCACTGGTAGGAGAAATGTTGGTACATCATCAAGATCTACCTTTA 60
QY 255 AACGAGAGAGAGAGAAAGACGAGCTATGCGACCTAGATCCTTA 300
DB 61 AACGAGAGAGAGAGAAAGACGAGCTATGCGACCTAGATCCTTA 106

RESULT 11
AAI57541
ID AAI57541 standard; DNA; 467 BP.
XX
AC AAI57541;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #26227 used to measure gene expression in human placenta sample.
XX

```

```

KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-48897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 26227; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;

Query Match 35.3%; Score 106; DB 22; Length 467;
Best Local Similarity 100.0%; Pred. No. 9.1e-42;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 AGTATGCTCAGCCACTGGTAGGAGAAATGTTGGTACATCATCAAGATCTACCTTTA 254
DB 1 AGTATGCTCAGCCACTGGTAGGAGAAATGTTGGTACATCATCAAGATCTACCTTTA 60
QY 255 AACGAGAGAGAGAGAAAGACGAGCTATGCGACCTAGATCCTTA 300
DB 61 AACGAGAGAGAGAGAAAGACGAGCTATGCGACCTAGATCCTTA 106

RESULT 12
AAD35994
ID AAD35994 standard; DNA; 1907 BP.
XX
AC AAD35994;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy2 DNA.
XX
KW Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;

```


CC and organ growth and development (e.g. the heart). The nervous system
CC disorders include lesions of central or peripheral nervous systems,
CC including traumatic lesions, ischaemic lesions, infectious lesions,
CC degenerative lesions, lesions associated with nutritional diseases or
CC disorders, neurological lesions, and lesions caused by toxic substances.
CC The neuropilin-like proteins and cDNAs are also useful as markers for
CC cancers. The neuropilin-like proteins are useful for regulating cell
CC proliferation, cell differentiation, stem cell growth factor activity,
CC for inducing proliferation of neural cells, regeneration of nerve and
CC brain tissue, for treatment of central and peripheral nervous system
CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopaenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues and as a food supplement or molecular
CC weight marker. The cDNAs are useful in gene identification, genome
CC mapping, transgenics, as hybridisation probes, for primer design, for
CC gene chips and as a DNA antigen. The present sequence encodes
CC neuropilin-Hy2.
XX
SQ Sequence 1907 BP; 548 A; 449 C; 406 G; 504 T; 0 other;
Query Match 19.3%; Score 58; DB 24; Length 1907;
Best Local Similarity 100.0%; Pred. No. 2.1e-18;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AACAGAGAAGAAAAAACTGAAGCACCTATGACTTACCTTACTGGGACCGGCAGGT 58
|||||
Db 1631 AACAGAGAAGAAAAAACTGAAGCACCTATGACTTACCTTACTGGGACCGGCAGGT 1688
RESULT 14
ABN59799
ID ABN59799 standard; cDNA: 1907 BP.
XX
AC ABN59799;
XX
DT 28-JUN-2002 (first entry)
XX
DE Novel human coding sequence SEQ ID NO: 210.
XX
KW Human; antinaemic; vulnary; antinflammatory; immunomodulator;
KW antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US26015.
XX
PR 11-SEP-2000; 2000US-0659671.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-292408/33.
DR P-PSDB; ABB97386.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis -
XX
PS Claim 1; SEQ ID NO 210; 509pp; English.
XX
CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate

CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a coding sequence of the
CC invention.
XX
SQ Sequence 1907 BP; 548 A; 449 C; 406 G; 504 T; 0 other;
Query Match 19.3%; Score 58; DB 24; Length 1907;
Best Local Similarity 100.0%; Pred. No. 2.1e-18;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AACAGAGAAGAAAAAACTGAAGCACCTATGACTTACCTTACTGGGACCGGCAGGT 58
|||||
Db 1631 AACAGAGAAGAAAAAACTGAAGCACCTATGACTTACCTTACTGGGACCGGCAGGT 1688
RESULT 15
AAH34855
ID AAH34855 standard; cDNA: 2428 BP.
XX
AC AAH34855;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1937.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR P-PSDB; AAG75450.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 3444-3445; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX

SQ Sequence 2428 BP; 677 A; 560 C; 535 G; 650 T; 6 other;

Query Match 19.3%; Score 58; DB 22; Length 2428;

Best Local Similarity 100.0%; Pred. No. 2.le-18;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACAGAAAGAAAAAACTGAAGGCACCTATGACTTACTGGGACCGGGCAGGT 58

|||||

Db 1601 AAACAGAAAGAAAAAACTGAAGGCACCTATGACTTACTGGGACCGGGCAGGT 1658

Search completed: January 21, 2003, 13:49:53

Job time : 179.276 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 10:13:26 ; Search time 160.669 Seconds
(without alignments)
3854.498 Million cell updates/sec

Title: US-10-060-830-6

Perfect score: 275

Sequence: 1 caacttcagttggtcagccc.....tttgaagatgatgtgctttt 275

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

	1:	2:	3:	4:	5:	6:	7:	8:	9:	10:	11:	12:	13:	14:	15:	16:	17:	18:	19:	20:	21:	22:	23:	24:
	/SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*	/SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	580	22	ABA60693 Human foetal liver
2	275	100.0	580	22	AAK08975 Human brain expres
3	275	100.0	580	22	AAK34866 Human bone marrow
4	275	100.0	580	22	AAI40582 Probe #9268 used t
5	275	100.0	2046	21	AAZ51872 Human Factor 8 Hom
6	260	94.5	594	23	AAK55485 DNA encoding novel
7	257	93.5	467	22	ABA76819 Human foetal liver
8	257	93.5	467	22	AAK25439 Human brain expres
9	257	93.5	467	22	AAK51451 Human bone marrow

10	257	93.5	467	22	AAI57541	Probe #26227 used
11	51	18.5	51	22	AAI73442	Human silent SNP c
12	25	9.1	51	22	AAI73443	Human silent SNP c
13	19	6.9	3751	23	ABL19472	Drosophila melanog
14	19	6.9	9417	22	AAK89089	Human digestive sy
15	19	6.9	9417	22	AAI57634	Human colorectal c
16	19	6.9	20892	23	ABL22310	Drosophila melanog
17	19	6.9	240825	22	AAF24497	Human PC-3 gene.
18	18	6.5	390	22	AAK40234	DNA encoding human
19	18	6.5	390	22	AAI02149	Human reproductive
20	18	6.5	399	22	AAI91187	Human polynucleoti
21	18	6.5	1286	24	ABK71686	Human dithp polynu
22	18	6.5	1686	22	AAK02811	HMEIR04 cDNA clone
23	18	6.5	1747	22	AAI59169	Human polynucleoti
24	18	6.5	2128	19	AAZ96356	S. pneumoniae deri
25	18	6.5	2128	19	AAV43007	Streptococcus pneu
26	18	6.5	2700	22	AAK02810	HFICU08 cDNA clone
27	18	6.5	3510	24	ABQ90026	M. capsulatus gene
28	18	6.5	9377	24	ABA97530	Human aminotransfe
29	18	6.5	26385	19	AAV52136	Streptococcus pneu
30	17	6.2	154	21	AAK17776	Human secreted pro
31	17	6.2	155	21	AAC08402	Human secreted pro
32	17	6.2	229	24	ABA9281	Cocoa husk-specifi
33	17	6.2	390	22	AAK23178	DNA encoding novel
34	17	6.2	460	24	ABN94039	Gene #537 used to
35	17	6.2	481	22	AAK64675	Human immune/haema
36	17	6.2	527	24	ABNG2417	Human cancer relat
37	17	6.2	577	22	ABA61003	Human foetal liver
38	17	6.2	577	22	AAK09295	Human brain expres
39	17	6.2	577	22	AAK35184	Human bone marrow
40	17	6.2	577	22	AAI40900	Probe #9586 used t
41	17	6.2	577	24	ABO9618	Human genome-deriv
42	17	6.2	613	24	ABO66278	Arabidopsis thalia
43	17	6.2	645	23	AAK64676	DNA encoding novel
44	17	6.2	647	21	AAK35198	Arabidopsis thalia
45	17	6.2	693	21	AAA43917	Human secreted exp

ALIGNMENTS

RESULT 1

ID	ABA60693	standard; DNA; 580 BP.
XX	ABA60693;	
XX	01-FEB-2002 (first entry)	
DT	Human foetal liver single exon nucleic acid probe #8998.	
DE	Human; foetal liver; gene expression; single exon nucleic acid probe; ss.	
KW	Homo sapiens.	
OS	WO200157277-A2.	
PN	09-AUG-2001.	
XX	30-JAN-2001; 2001WO-US000669.	
PF	04-FEB-2000; 2000US-0180312.	
PR	26-MAY-2000; 2000US-0207456.	
PR	30-JUN-2000; 2000US-0608408.	
PR	03-AUG-2000; 2000US-0632366.	
PR	21-SEP-2000; 2000US-0234687.	
PR	27-SEP-2000; 2000US-0236359.	
PR	04-OCT-2000; 2000GB-0024263.	
XX	(MOLE-) MOLECULAR DYNAMICS INC.	
PA	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX		

```
DR WPI; 2001-483447/52.
XX PA
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PS
XX PS Claim 1: SEQ ID NO 8998; 639pp + sequence listing; English.
XX CC
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human fetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC fetal liver. The present sequence is a single exon nucleic acid
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;
    Query Match 100.0%; Score 275; DB 22; Length 580;
    Best Local Similarity 100.0%; Pred. No. 2.8e-126;
    Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAACTTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAACTCTCCCC 60
DB 281 CAACTTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAACTCTCCCC 340
QY 61 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGTGTCTCTCAGCCGAGG 120
DB 341 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGTGTCTCTCAGCCGAGG 400
QY 121 CCCAGTATGATACCCCGAAAGCTGGGAGCCAGGTCTACCTGCCCGAGCAAGTGGTGT 180
DB 401 CCCAGTATGATACCCCGAAAGCTGGGAGCCAGGTCTACCTGCCCGAGCAAGTGGTGT 460
QY 181 ACCAGGTGCCACAGACACACAAGAGTATCAGGAGCAGGAAGGATGGGGAATGTGATG 240
DB 461 ACCAGGTGCCACAGACACACAAGAGTATCAGGAGCAGGAAGGATGGGGAATGTGATG 520
QY 241 TTTTAAAGAAGTCTTTGAAGATGATGCTGCTTT 275
DB 521 TTTTAAAGAAGTCTTTGAAGATGATGCTGCTTT 555
RESULT 2
AAK08975
ID AAK08975 standard; DNA; 580 BP.
AC AAK08975;
XX
XX 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe SEQ ID NO: 8996.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0623266.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0234687.
XX PR 04-OCT-2000; 2000GB-0024263.
```

```
XX PA
XX PI (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX PS
XX PS Example 4; SEQ ID NO: 8966; 650pp + sequence listing; English.
XX CC
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is one of the probes of the
XX CC invention.
XX SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;
    Query Match 100.0%; Score 275; DB 22; Length 580;
    Best Local Similarity 100.0%; Pred. No. 2.8e-126;
    Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAACTTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAACTCTCCCC 60
DB 281 CAACTTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAACTCTCCCC 340
QY 61 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGTGTCTCTCAGCCGAGG 120
DB 341 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGTGTCTCTCAGCCGAGG 400
QY 121 CCCAGTATGATACCCCGAAAGCTGGGAGCCAGGTCTACCTGCCCGAGCAAGTGGTGT 180
DB 401 CCCAGTATGATACCCCGAAAGCTGGGAGCCAGGTCTACCTGCCCGAGCAAGTGGTGT 460
QY 181 ACCAGGTGCCACAGACACACAAGAGTATCAGGAGCAGGAAGGATGGGGAATGTGATG 240
DB 461 ACCAGGTGCCACAGACACACAAGAGTATCAGGAGCAGGAAGGATGGGGAATGTGATG 520
QY 241 TTTTAAAGAAGTCTTTGAAGATGATGCTGCTTT 275
DB 521 TTTTAAAGAAGTCTTTGAAGATGATGCTGCTTT 555
RESULT 3
AAK34866
ID AAK34866 standard; DNA; 580 BP.
XX
XX AC AAK34866;
XX
XX 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 9423.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0623266.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0234687.
XX PR 04-OCT-2000; 2000GB-0024263.
```

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 9423; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.

XX Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;

Query Match 100.0%; Score 275; DB 22; Length 580;

Best Local Similarity 100.0%; Pred. No. 2.8e-126;

Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAACCTCCCC 60

Db 281 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAACCTCCCC 340

Qy 61 CACTAGTGGGAACCTTACAATACACTTCTCCAGGACTGACAGCTGCTCCTCAGCCCAAGG 120

Db 341 CACTAGTGGGAACCTTACAATACACTTCTCCAGGACTGACAGCTGCTCCTCAGCCCAAGG 400

Qy 121 CCCAGTATGATACCCCGAAAGCTGGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGTGT 180

Db 401 CCCAGTATGATACCCCGAAAGCTGGGAAGCTGGAAGCTGGAAGCTGGAAGTGT 460

Qy 181 ACCAGTCCACAGACGACACAAAGTATCAGGAGCAGGAGGATGGGGAATGTGATG 240

Db 461 ACCAGTCCACAGACGACACAAAGTATCAGGAGCAGGAGGATGGGGAATGTGATG 520

Qy 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275

Db 521 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 555

RESULT 4

AAI40582

ID AAI40582 standard; DNA; 580 BP.

XX AC

XX AAI40582;

XX DT

XX 17-OCT-2001 (first entry)

XX DE Probe #9268 used to measure gene expression in human placenta sample.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX KW genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -

XX Claim 25; SEQ ID No 9268; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.

XX Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;

Query Match 100.0%; Score 275; DB 22; Length 580;

Best Local Similarity 100.0%; Pred. No. 2.8e-126;

Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAACCTCCCC 60

Db 281 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCAACCTCCCC 340

Qy 61 CACTAGTGGGAACCTTACAATACACTTCTCCAGGACTGACAGCTGCTCCTCAGCCCAAGG 120

Db 341 CACTAGTGGGAACCTTACAATACACTTCTCCAGGACTGACAGCTGCTCCTCAGCCCAAGG 400

Qy 121 CCCAGTATGATACCCCGAAAGCTGGGAAGCTGGAAGCTGGAAGCTGGAAGTGT 180

Db 401 CCCAGTATGATACCCCGAAAGCTGGGAAGCTGGAAGCTGGAAGCTGGAAGTGT 460

Qy 181 ACCAGTCCACAGACGACACAAAGTATCAGGAGCAGGAGGATGGGGAATGTGATG 240

Db 461 ACCAGTCCACAGACGACACAAAGTATCAGGAGCAGGAGGATGGGGAATGTGATG 520

Qy 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275

Db 521 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 555

RESULT 5

AAZ51872

ID AAZ51872 standard; cDNA; 2046 BP.

XX AC

XX AAZ51872;

XX DT

XX 04-JUL-2000 (first entry)

XX DE Human Factor 8 Homologue cDNA.

XX KW Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;

XX KW cerebroprotective; therapeutic; coagulation related disorder;

XX KW haemophilia; stroke; screening; ss.

XX OS Homo sapiens.

XX PN Key

XX FT CDS

XX FT Location/Qualifiers

XX FT 19..2028

XX FT /tag= a

XX FT /product= "Factor 8 Homologue"

XX PN WO200012532-A1.

XX PD 09-MAR-2000.

XX XX

PF 20-AUG-1999; 99WO-US19047.
XX
PR 31-AUG-1998; 98US-0098521.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Rostek PRJ, Su W, Li XM;
XX
XX WPI; 2000-256580/22.
DR P-PSDB; AAY70539.
XX
XX
PT Factor 8 homolog polypeptides and nucleic acids encoding them for
PT treating coagulation related disorders such as hemophilia and stroke
XX
XX
PS Claim 1; Page 61-64; 68pp; English.
XX
XX The present sequence is a cDNA encoding human Factor 8 homologue (F8H)
CC which is a coagulation cofactor. The protein is selectively expressed in
CC haematopoietic, heart and reproductive tissues. It has haemostatic and
CC cerebroprotective activities. The F8H contains a Factor 5/8 signature
CC and is useful as a therapeutic for treating coagulation related diseases
CC such as haemophilia and stroke. The nucleic acid is useful as
CC hybridisation probe and amplification primer for detecting deficiencies
CC in the level of F8H mRNA, for screening F8H gene mutations and for
CC monitoring regulation of gene expression. Fragments of the nucleic acid
CC are also useful as diagnostic probes and primers, and can be used in
CC screening methods such as those using DNA chips. The present sequence is
CC also useful as a target to screen therapeutically useful modulators
CC of the F8H.
XX
XX Sequence 2046 BP; 615 A; 457 C; 467 G; 507 T; 0 other;
XX
Query Match 100.0%; Score 275; DB 21; Length 2046;
Best Local Similarity 100.0%; Pred. No. 2.8e-126;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTCAGGCTACGGGACCAACTCCCC 60
Db 1769 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTCAGGCTACGGGACCAACTCCCC 1828
QY 61 CACTAGTGGGAACTTACATACTCTCTCCAGGACTGACAGCTGCTCTCCAGCCAGG 120
Db 1829 CACTAGTGGGAACTTACATACTCTCTCCAGGACTGACAGCTGCTCTCCAGCCAGG 1888
QY 121 CCCAGTATGATACCCCGAAGCTGGGAAGCCAGGTCTACCTGCCCCAGCAATGGTGT 180
Db 1889 CCCAGTATGATACCCCGAAGCTGGGAAGCCAGGTCTACCTGCCCCAGCAATGGTGT 1948
QY 181 ACCAGGTCCACAGACACACAAAGTATCAGGACGAGGAGGATGGGAAATGTGATG 240
Db 1949 ACCAGGTCCACAGACACACAAAGTATCAGGACGAGGAGGATGGGAAATGTGATG 2008
QY 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275
Db 2009 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 2043
RESULT 6
AAS65485
ID AAS65485 standard; cDNA; 594 BP.
XX
AC AAS65485;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #1289.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.

XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG01298.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID NO 1289; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 594 BP; 175 A; 160 C; 139 G; 120 T; 0 other;
XX
Query Match 94.5%; Score 260; DB 23; Length 594;
Best Local Similarity 100.0%; Pred. No. 7.5e-119;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTTCAAGGCTACGGGACCAACTCCCC 60
Db 335 CAACCTCAGTTGGTCCAGCCCTCCACATCCACTTTTCAAGGCTACGGGACCAACTCCCC 394
QY 61 CACTAGTGGGAACTTACATACTCTCTCCAGGACTGACAGCTGCTCTCCAGCCAGG 120
Db 395 CACTAGTGGGAACTTACATACTCTCTCCAGGACTGACAGCTGCTCTCCAGCCAGG 454
QY 121 CCCAGTATGATACCCCGAAGCTGGGAAGCCAGGTCTACCTGCCCCAGCAATGGTGT 180
Db 455 CCCAGTATGATACCCCGAAGCTGGGAAGCCAGGTCTACCTGCCCCAGCAATGGTGT 514
QY 181 ACCAGGTCCACAGACACACAAAGTATCAGGACGAGGAGGATGGGAATGTGATG 240
Db 515 ACCAGGTCCACAGACACACAAAGTATCAGGACGAGGAGGATGGGAATGTGATG 574
QY 241 TTTTAAAGAAATCCTTTGA 260
Db 575 TTTTAAAGAAATCCTTTGA 594
RESULT 7
ABA76819

```
ID ABA76819 standard; DNA: 467 BP.
XX
AC ABA76819;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #25124.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver.
XX
XX Claim 4; SEQ ID NO 25124; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;
XX
XX Query Match 93.5%; Score 257; DB 22; Length 467;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-117;
XX Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAACCTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCACTCCCC 60
DB 211 CAACCTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCACTCCCC 270
QY 61 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCCTCAGCCCCAGG 120
DB 271 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCCTCAGCCCCAGG 330
QY 121 CCCAGTATGATACCCCGAAAGCTGGGAAGCTGGAAGCTGCTACCTCCCCAGACGAATGGTGT 180
DB 331 CCCAGTATGATACCCCGAAAGCTGGGAAGCTGGAAGCTGCTACCTCCCCAGACGAATGGTGT 390
QY 181 ACCAGTGCCACAGAGCACACAAGAAGTATCAGGAGCAGGAAGGATGGGAATGTGTATG 240
DB 391 ACCAGTGCCACAGAGCACACAAGAAGTATCAGGAGCAGGAAGGATGGGAATGTGTATG 450
QY 241 TTTTAAAGAAATCCTT 257
DB 451 TTTTAAAGAAATCCTT 467
```

```
RESULT 8
AAK25439
ID AAK25439 standard; DNA: 467 BP.
XX
XX AAK25439;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 25430.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO: 25430; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;
XX
XX Query Match 93.5%; Score 257; DB 22; Length 467;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-117;
XX Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAACCTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCACTCCCC 60
DB 211 CAACCTTCAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCACTCCCC 270
QY 61 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCCTCAGCCCCAGG 120
DB 271 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCCTCAGCCCCAGG 330
QY 121 CCCAGTATGATACCCCGAAAGCTGGGAAGCTGGAAGCTGCTACCTCCCCAGACGAATGGTGT 180
DB 331 CCCAGTATGATACCCCGAAAGCTGGGAAGCTGGAAGCTGCTACCTCCCCAGACGAATGGTGT 390
QY 181 ACCAGTGCCACAGAGCACACAAGAAGTATCAGGAGCAGGAAGGATGGGAATGTGTATG 240
DB 391 ACCAGTGCCACAGAGCACACAAGAAGTATCAGGAGCAGGAAGGATGGGAATGTGTATG 450
QY 241 TTTTAAAGAAATCCTT 257
```

```
QY 1909 CCAGGCGAGAGTTTATCATGCTATGCTGAACCACTCCCAATACGGGGCCTGAGTAT 1968
Db 653 PrometGlnGluValTyrHisAlaTyrAlaGluProLeuProValThrGlyProGluTyr 672
QY 1969 GCAACCCCATCATGACATGATGTCAGGCGCCACCACTTCAGTTGGTCAGCCCTCC 2028
Db 673 AlaThrProIleValMetAspMetSerGlyHisProThrAlaSerValGlyLeuProSer 692
QY 2029 ACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCTCCCTAGCTAGTGGAACTTACAATACA 2088
Db 693 ThrSerThrPhelystThrAlaGlyThrGlnProHisAlaLeuValGlyThrThrThrAsnThr 712
QY 2089 CTTCTCTCAGGACTGACAGCTGCTCCTCAGCCAGCCAGCATGATACCCGAAAGCT 2148
Db 713 LeuLeuSerArgThrAspSerCysSerGlyGlnAlaGlnTyrAspThrProLysGly 732
QY 2149 GGAAGCCAGGCTACCTCCCTCCCTCCCTCAGCCAGCCAGGTCGACAGTGCCACAGACACAA 2208
Db 733 GlyLys---SerAlaAlaThrProGluGluLeuValTyrGlnValProGlnSerThrGln 751
QY 2209 GAAGTATCAGGAGGAGGAGGAGGAGTGTGATGTTTAAAGAAATCCTT 2262
Db 752 GluLeuSerGlyAlaGlyArgAspGluLysPheAspAlaPheLysGluIleLeu 769
RESULT 5
Q14089.
ID Q14089 PRELIMINARY; PRT; 364 AA.
AC Q14089;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 40.0 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shibata T.;
RT "unpublished.";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; D29810; BAA18909.1; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;

Alignment Scores:
Pred. No.: 1.59e-129 Length: 364
Score: 1625.00 Matches: 342
Percent Similarity: 94.81% Conservative: 5
Best Local Similarity: 93.44% Mismatches: 14
Query Match: 39.22% Indels: 11
Db: 4 Gaps: 2

US-10-060-830-1 (1-2280) x Q14089 (1-364)

QY 145 GATGATGGACACACTGTACTAGCCCTGAGAGTGGAACTTACATCCATAAATAC 204
Db 1 AspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerIleAsnTyr 20
QY 205 CCACAGACCTATCCCAACAGCACTGTTGTGAATGGGAGATCCGTGTAAGATGGGAGAG 264
Db 21 ProGlnThrTyrProAsnSerThrValCysGluTyrGluIleArgValLysMetGlyGlu 40
```

```
QY 265 AGATTTCGCATCAAAATTTGGTGACTTTGACATTGAAGATTCTGATTCTTGTCACATTAAT 324
Db 41 ArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCysHisPheAsn 60
QY 325 TACTTGAGAAATTTAATGAATTTGGAGTCAGCAGAACTGAAATAGCAAAATCTGTGTT 384
Db 61 TyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLysTyrCysGly 80
QY 385 CTGGGGTTCCAAATGAACCATTCATTAATCAAAAGGCAATGAATCACAATTCATGCTGTTTC 444
Db 81 LeuGlyLeuGlnIleAsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPhe 100
QY 445 ATGAGTGGAAATCCATGTTCTGGACGCGGATTTTGGCCCTCATCTCTGTATAGATAAA 504
Db 101 MetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLys 120
QY 505 CA-AGATCAATTAATTAATTTGGACACTGCATCCAAAT- TTTTGGAACTGAGTTCCAGTA 562
Db 121 GlnArgSerAsnTyrLeuPheGlyHisCysIleGlnPhePheTyrAsnLeuSerSerVal 140
QY 563 AGTACTGC-CCAGCTGTTGCTCTGCTTCTTCTGCTGAGATATCTGGAACAATTCCTCAT 621
Db 141 SerThrAlaProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIleProHis 160
QY 622 GGATATAGAGATTCCTCGCCATTTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 681
Db 161 GlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValValSerAsn 180
QY 682 AGTTGGGGCGGCAATCATAGTGTGTAATAGTAAGTATTCCTCTATTTATGAAAGTTCT 741
Db 181 ThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGluSerSer 200
QY 742 TTGGCTAAACAACGTCACATCTGTGGGGGACACTATCTACAAGTCTT- -TTTACATTT 798
Db 201 LeuAlaAsnValThrSerValValGlyHisLeuIleLeuGlnValPhePheThrPhe 220
QY 799 AAGACAAGTGGATGTTTATGAACACTGGGGATGGAGTCTGGTGT- GATCGCGATCTCA 857
Db 221 LysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyGlyAspArgGlySerSer 240
QY 858 AATACAGCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 917
Db 241 AsnAsnSerIle- ThrValLeuGluTyrThrAspHisThrGlyGlnGluAsnSerTyr 260
QY 918 ACCCAAAAAGC- CAGGCTGAAAACCTGGACCCGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 976
Db 260 SProlysLysSerGlnAlaGluLysThrTrpThrAlaLeuGly- AlaPheAlaThrAspG 280
QY 977 AATACAGTGGTTACAATAGATTGTAATAAGAAAAGAAAATAACAGGCATTTATAACCA 1036
Db 280 LutyGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThrT 300
QY 1037 CTGGATCCACCATGGTGGAGCAAAATTAATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1096
Db 300 hrGlySerThrMetVal- SerThrIleThrMetCysLeuProThrGluSerCysThrVal 319
QY 1097 ATGATGGGCGAGAAATGGAGTGTGACAGAGACCTGGT- GTGAGCAAGATTAAGATATTT 1155
Db 320 MetMetGlyArgAsnGlyLeuCysThrGluSerLeuValValGluGlnAspLysIlePhe 339
QY 1156 CAAGGAACAAAGAT- -TATCACCAGGATGTCGCTGAATAACTTTTGGCCACCAATTTAT 1212
Db 340 GlnGlyAsnLysArgIleIleThrArgMetValArgAsnAsnPheLeuProIleIle 359
QY 1213 GCACGTTTATT 1224
Db 360 AlaArgLeuLeu 363
RESULT 6
Q9D9K5
ID Q9D9K5 PRELIMINARY; PRT; 251 AA.
AC Q9D9K5;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
```

DT	01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT	01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE	1700055P21Rik protein.
GN	ESDN OR 1700055P21Rik.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX	NCBI_TaxId=10090;
RP	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;
RX	MEDLINE=21085660; PubMed=11217851;
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA	Arakawa T., Hara A., Nishikishi I., Konno H., Adachi J., Fukuda S.,
RA	Alzawa K., Izawa M., Fukui K., Kiyosawa H., Kondo S., Yamanaka I.,
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA	Hayashizaki Y.;
RT	"Functional annotation of a full-length mouse cDNA collection.";
RL	Nature 409:685-690(2001).
DR	EMBL; AK006805; BAB24750.1; .
DR	MGD; MGI:1920629; ESDN.
SQ	SEQUENCE 251 AA; 27428 MW; F86E0AA15EF51AE9 CRC64;

Alignment Scores:			
Pred. No.:	2,776-89	Length:	251
Score:	1151.50	Matches:	218
Percent Similarity:	91.16%	Conservative:	21
Best Local Similarity:	87.55%	Mismatches:	29
Query Match:	27.79%	Indels:	1
DB:	11	Gaps:	1

US-10-060-830-1 (1-2280) x Q9D9K5 (1-251)	
QY 1516 GTAGCGCTGCCTGCAGTCTTGTGCCCTGTGCTGGTCATGCTCCTCACTACTCTCATTCFC	1575
Db 4 ValAlaLeuAlaAlaValLeuValProValLeuValMetAlaLeuThrThrLeuLeuLeu	23
QY 1576 ATATTAGTGTGCTTGGCAGTGGAGAAACAGAAAGAAACCTGAGGACCTATGAC	1635
Db 24 IleLeuValCysAlaIrrPHisTrrPArgNArgLysLysThrGlulGlyAlaIrrAsp	43
QY 1536 TTACCTTACTGGACCGGCGAGTGGTGGAAAGGAATGAAGCAGTCTTCTCTCGCAAAA	1695
Db 44 LeuProHisTrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnLeuLeuProAlaLys	63
QY 1696 GCAGTGACCATGAGGAACCCAGTTCGGCTATAGCAGCGAAGTAAATCACTACCTGAGT	1755
Db 64 SerValAspHisGluGluThrProValIrrtyrSerThrSerGluValSerHisLeuSer	83
QY 1756 CCAAGAGAAGTCACACAGTGCCTGCAGGCTGACTCTGCAGAGTATGCTCAGCACCTGGTA	1815
Db 84 AlaArgGluValThrThrValLeuGlnAlaAspSerAlaGluIrrAlaGlnProLeuVal	103
QY 1816 GGAGGAATGTGTGTACACTTCATCAAGATCTACCTTTAAACCAGAGAAGGAAGAA	1875
Db 104 GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGlu	123
QY 1876 GCAGCTATGCAGACCTAGATCCTTACAACTCACCGGCGAGGAAGTTTATCATGCGCTAT	1935
Db 124 AlaGlyTrrAlaAspLeuAspProTrrAsnSerPrometGlnGluValIrrIrrIrrIrr	143

QY 1936 GCTGAACCACTCCCAATTACGGGGCCCTAGTATGCACACCCCAATCATCATGGACATCTCA 1995
 Db 144 AlaGluProLeuProValThrGlyProGluTyrAlaThrProIleValMetaspMeSer 163
 QY 1996 GGGCACCCCAACACTTCAGTTGGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAAC 2055
 Db 164 GlyHisProThrAlaSerValGlyLeuProSerThrSerThrPheLysThrAlaGlyThr 183
 QY 2056 CAACCTCCCCACACTAGTGGGAACCTTACAAATACACTTCTCTCCAGACTGACAGGTGCTCC 2115
 Db 184 GlnProHisAlaLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSer 203
 QY 2116 TCAGCCAGGCGCCAGTATGATACCCGGAAGCTGGGAAGCCAGGCTCTACCTGCCCCAGAC 2175
 Db 204 SerGlyGlnAlaGlnTyrAspThrProLysGlyGlyLys---SerAlaAlaThrProGlu 222
 QY 2176 GAATGGTGTACCCAGGTCCAGAGCACACAGAAGTATCAGGAGCAGGAGGATGGG 2235
 Db 223 GluLeuValTyrGlnValProGlnSerThrGlnGluLeuSerGlyAlaGlyArgAspGlu 242
 QY 2236 GAATGTGATGTTTAAAGAAATCCTT 2262
 Db 243 LysPheAspAlaPheLysGluLeuLeu 251
 RESULT 7
 Q9DAJ3 PRELIMINARY: PRT: 503 AA.
 AC Q9DAJ3
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 4631413K1IRik protein.
 GN 4631413K1IRIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadoita K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schraml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL; AK016485; BAB30265.1; -;
 DR MGD; MGI:1913936; 4631413K1IRik.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR004043; LCCL_dom.
 DR Pfam; PF00431; CUB; 1.
 DR SMART; SM00042; CUB; 1.
 DR PROSITE; PS01180; CUB; 1.
 SQ SEQUENCE 503 AA; 54547 MW; FEB121E845CAA06B CRC64;
 Alignment Scores:

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 21, 2003, 09:58:26 ; Search time 5065.02 seconds
(without alignments)
4188.719 Million cell updates/sec

Title: US-10-060-830-3
Perfect score: 3888
Sequence: 1 MPFLLLLLLLLLLEADG.....TQEVSGAGRGDCGVFKEIL 729

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q/cygn2_1/USPRO.spool/US10060830/runat_16012003_091559_29008/app_query.fasta_1.1742
-DB=GenEmbl -Qfmt=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830 -CGN1_1_6337 -runat_16012003_091559_29008 -NCPU=6 -ICPU=3
-NOLXPY -NOLMAP -LARGESUBRY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_un.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rnd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3888	100.0	2939	9	AF387547 Homo sapi
2	3888	100.0	6093	9	AB073146 Homo sapi
3	3871	99.6	3858	9	BC029658 Homo sapi
4	3415.5	87.8	2339	10	AF387549 Rattus no
5	3382.5	87.0	2642	10	AF387548 Mus muscu
6	2692	69.2	2020	9	AK056350 Homo sapi
7	2592.5	66.7	1907	6	AX405795 Sequence
8	2124.5	54.6	1388	9	HUMCUB1 D29810 Human mRNA
9	1033.5	26.6	2547	9	AK095973 Homo sapi
10	924.5	23.8	1620	6	AX118820 Sequence
11	924.5	23.8	1761	6	AX118818 Sequence
12	922.5	23.7	192247	2	AC013497 Homo sapi
13	922.5	23.7	193623	9	AC106728 Homo sapi
14	916.5	23.6	1768	6	AX118822 Sequence
15	860.5	22.1	1464	6	AX118816 Sequence
16	852.5	21.9	203102	2	AC126975 Rattus no
17	825	21.2	145515	2	AC025661 Homo sapi
18	681.5	17.5	586	9	BC007117 Homo sapi
19	442.5	11.4	164216	2	AC091213 Homo sapi
20	429	11.0	192247	2	AC013497 Homo sapi
21	428.5	11.0	2781	10	AF483506 Mus muscu
22	428.5	11.0	2781	10	AF483507 Mus muscu
23	425.5	10.9	4765	10	AF022857 Mus muscu
24	425.5	10.9	4780	10	AF022858 Mus muscu
25	423.5	10.9	4718	10	AF022854 Mus muscu
26	423.5	10.9	4733	10	AF022861 Mus muscu
27	423.5	10.9	4769	10	AF022855 Mus muscu
28	423.5	10.9	4784	10	AF022856 Mus muscu
29	423	10.8	3371	10	AF016297 Rattus no
30	419.5	10.8	2670	6	ES9779 Soluble neu
31	412	10.6	2846	10	BC026771 Mus muscu
32	411	10.6	2730	9	AF022859 Homo sapi
33	411	10.6	2781	9	AF022860 Homo sapi
34	411	10.6	2796	9	AF016098 Homo sapi
35	411	10.6	3357	9	AF280544 Homo sapi
36	411	10.6	3372	9	AF280545 Homo sapi
37	410	10.5	2760	5	AF417235 Gallus ga
38	410	10.5	2811	5	AF417236 Gallus ga
39	400.5	10.3	1922	6	AX147496 Sequence
40	400.5	10.3	1972	9	AX145712 Homo sapi
41	400.5	10.3	2014	9	BC007737 Homo sapi
42	400.5	10.3	2232	9	BC007533 Homo sapi
43	400.5	10.3	2241	6	AX147494 Sequence
44	400.5	10.3	2568	6	AX133545 Sequence
45	400.5	10.3	2772	9	AF016050 Homo sapi

ALIGNMENTS

[illegible]

BASE COUNT .	1671 a	1258 c	1289 g	1874 t	1 others
ORIGIN					

Alignment Scores:

Pred. No.:	2.21e-258	Length:	6093
Score:	3888.00	Matches:	729
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-060-830-3 (1-729) x AB073146 (1-6093)

QY 1 MetProLeuPheLeuLeuLeuValLeuLeuLeuLeuGluAspAlaGlyAla 20

Db 502 ATGCCCTCTGTTCCCTCCTGCTCTTACTTGTCCCTGCTCCTGCTGCTCGAGGACGCTGGAGCC 561

QY 21 GlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyValThrLeuThrSer 40

Db 562 CAGCAAGGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAACCCCTTACATCC 621

QY 41 IleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluIleArqValLys 60

Db 622 ATAACTACCCACAGACCTATCCCAACAGCACTGTTTGTGAATGGGAGATCCGTGTAAAG 681

QY 61 MetGlyGluArgValArgIleLeysPheGlyAspPheAspIleGluAspSerAspSerCys 80

Db 682 ATGGAGAGAGATTGCGATCAAATTGGTGACTTTGACATTGAAGATTCTGATTCTTGT 741

Qy 81 HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys 100

Db 742 CACTTTAACTTGAGAAATTATAATGGAA TTGGAGTCAGCAGAACTGAAATAGGCAA 801

QY 101 TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr 120

Db 802 TACTGTGGTCTGGGGTTGCAATGAACCATTC AATTGAATCAAAGGCAATGAAATCACA 861

Qy 121 LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyGlyPheLeuAlaSerTyrSerVal 140

Db 862 TTGCTGTT CATGAGTGGAATCCATGTTCTGGACGGGATTTTGGCCTCATACTCTGTT 921

Qy 141 IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu 160

Db 922 ATAGATAACAAGATCTAATTACTTGTGGACACTGCATCCAATTTTGGAACTGAG 981

QY 161 PheSerLysTyrCysProAlaGlyCysLeuLeuPropheAlaGluIleSerGlyThrIle 180

Db 982 TTCAGTAAGTACTGCCCAGCTGGTTGCTGCTGCTTTTCTCCTTTCTGAGATATCTGGAACAATT 104

Qy 181 ProHisGlyTyrArgAspSerProLeuCysMetAlaGlyValHisAlaGlyValVal 200

Db 1042 CCTCATGGATATAGAGATTCTCGCCATTGTCATGGCTGGTGCATGCAGGAGTAGTG 110

QY 201 SerAsnThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGlu 220

Db 1102 TCAACACGTTGGGCGCCAAATCAGTGTGTAATTAGTAAGGTATCCCCCTATTATGAA 116

Qy 221 SerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr 240

D_b 1162 AGTCTTTGGCTAACAAACGTCACATCTGTGGTGGGACACTTATCTACAAGCTTTTACA 122

Qy 241 PheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspPro 260

Db 1222 TTTAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATCGCGGATCCT 128

Qy 261 GlnIleThrAlaSerSerValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrp 280

Db 1282 CAAATAACAGCATCATCTGTGCTGGAGTGACTGACCACACAGGGCAAGAGAACAGTTGG 134

Qy 281 LysProLysLysAlaArgLeuLysLysProGlyProProTrpAlaAlaPheAlaThrAsp 300

Db 1342 AACCCAAAAGCCAGGCTGAAAAACCTGGACCGCCTTTGGCTGCTTTTGCCACTGAT 140

Qy	301	GluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysGlyIleThrGlyIleIleThr	320
Db	1402	GAATACCAGTGGTTACAAATAGATTTTGAATAGAGAAAGAAAATAACAGGCATTATACC	1461
Qy	321	ThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSer	340
Db	1462	ACTGGATCCCACTGGTGGAGCAACAATTACTATGTGTCTGCCTACAGAAATCCTGTACAGT	1521
Qy	341	AspAspGlyGlnLysTrpThrValTyrArgGluProGlyValcIuGlnAspLysIlePhe	360
Db	1522	GATCATGGCCAGAAATGGACTGTGTACAGAGAGCCGTGGTGTGGACAGATAAGATATTT	1581
Qy	361	GlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProPheIleLeuAla	380
Db	1582	CAAGAAACAAGATTATCACAGGATGTCGTATAAATCTTTTTCGCCACCAATATTGCA	1641
Qy	381	ArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeu	400
Db	1642	CGTTTTATTAGATGTAATCCTACCCAATGGCAGCAGAAAATTCGCCATGAAAATGGAGCTG	1701
Qy	401	LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProProPro	420
Db	1702	CTCGAGTGTCAAGTTATCTCAAGGTGTCTCTCCAAAATCTTACTCAACCTCCACCTCCT	1761
Qy	421	ArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla	440
Db	1762	CGGAACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAAATAACGCAAAAGTCGTGCC	1821
Qy	441	ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu	460
Db	1822	CCAAAATTTACGCAACCACTACAACTCCGAGTACGCAATGAAATTTCTGTCACAGACAGAA	1881
Qy	461	GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp	480
Db	1882	CAACAACCTGCCAGTCTCTGATATCAGAAATACTACCGTAACCTCAAAATGTACCAAAAGAT	1941
Qy	481	ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu	500
Db	1942	GTAGCGCTGGCTGCAGTTCVTGTCCCTGTGCTGGTTCATGGCTCCTCACTACTCTATTCTC	2001
Qy	501	IleLeuValCysAlaTrpHisTyrArgAsnArgLysLysLysThrGluClyThrTyrAsp	520
Db	2002	ATATTAGTGTGTCTGGCACTGGAGAAACAGAAAGAAAACCTGAAGGCACTTATGAC	2061
Qy	521	LeuProTyrTrpAspArgAlaGlyTyrTrpLysGlyMetLysGlnPheLeuProAlaLys	540
Db	2062	TTACCTTACTGGGACCGGCGAGTGTGGTGAAGGAATCAGACAGTTCTTCTCTGC AAA	2121
Qy	541	AlaValAspHisGluGluThrProValArgTyrSerSerSerGluValAsnHisLeuSer	560
Db	2122	GCAGTGGACCATAGGAAACCCCACTTCGTATAGCAGCAGCGAAAGTTAATCACCTGAGT	2181
Qy	561	ProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuVal	580
Db	2182	CCAAGAAAGTCCACACAGTGTCTCAGGCTGACTCTGCAGAGTATGTCTCAGCCACTGCTA	2241
Qy	581	GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGlu	600
Db	2242	GGAGGAATTTGGTGTACACTTCATCAAGAGTCTACCTTTAAACACAGAGAGGAAAGAA	2301
Qy	601	AlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyr	620
Db	2302	GCAGGCTATGCAGACCTAGATCCTTACAACCTCACAGGCGCAGGAAGTTTATCATGCCCTAT	2361
Qy	621	AlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetaspMetSer	640
Db	2362	GCTGAACCACTCCCAATTACGGGGCTGAGTATGCAACCCCAATCATCATGAGCATGTGCA	2421
Qy	641	GlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsn	660
Db	2422	GGGCACCCCAACTTCAGTTGTGTGCTAGCCCTCCACATCCACTTTCAAGGCTACGGGAAAC	2481
Qy	661	GlnProProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSer	680

BASE COUNT 612 a 606 c 594 q 527 t
KTCGRIRIKFGDFDIEDSDYCHLNYLKIIFNGICVSVRTETGKYCGGLGQMNQSTESKGS
EITVLVMSGHASGRFLASIVYSKODLITCLDTSVNSFLEPFESKYCAPGCLLPFAE
ISGITIPIHGRSDSPLCMAGIHCAGVSDGLGGQISVTSIKGTPYESLANNVTSWVGY
LSTSLFTTSGCYGLMGESVIAQDITASSVLEWTDHMGQENSKPEKABLRKPG
PWAFAATDEHOLDLNKEKKITGIVTSLIEHNYVYSAYRVLYSDDGQKWTVY
REFGAQDKIFQGNKDYHKDVNRNLFPIIAREIRVNPVOMQOKIAMKVELLCQFTL
KGBRLTOPPPRPNRSNNKTVTHPKLRAPKFTQALQPSRNDLPLTLPQAOTATP
VKNNTVTSYTDVLAALVYLVLMALTTILILICAWHRNRKKKAECTYDLPHWD
RAGWQKQVOLLPAKSVEHEETPVYSNESVSHLSPREVTVLQAADSAAEVAQPLVGSI
VGLTHQBSFTKPEEGKEASADLDPNYNAPQEVYHAYAEPLPVTGPXYATPIVMDMSG
HSTASVGLPSTSTFTFAGNQPPALDVTYNPVLSTSCSGGQAYQITPKGKFAAAPE
ELVYVQPSQTEASAGRDEKDFAKFTL"

QY 227 ValThrSerValValGlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCys 246
 Db 1121 GTCACTTCCACGGTGGGATATCTATCTGCAAGTCTGTTTACATTTAAGACAAGTGGTGC 1180
 QY 247 TyrGlyThrLeuGlyMetGlySerGlyValIleAlaAspProGlnIleThrAlaSerSer 266
 Db 1181 TATGGGACTCTGGGATGGAGTCTGGTGTGATTGCCGATCCCGAGATAAGCAGCATCGTCT 1240
 QY 267 ValLeuGlnThrAspHisThrGlyGlnLeuAsnSerTrpLysProLysLysAlaArg 286
 Db 1241 GCACGTGGAGTGGACTGACACATGGGGCAGAGACAGCTGGACAGCGGAGAGCCGAGG 1300
 QY 287 LeuLysLysProGlyProProTrpAlaAlaPheAlaThrAspGluTrpGlnTrpLeuGln 306
 Db 1301 CTGAGAAACCCGGGCTCCCTGGCTGCTTTTCCCACTGATGAGCATCACTGGCTGCAG 1360
 QY 307 IleAspLeuAsnLysGluLysLysIleThrGlyIleIleThrThrGlySerThrMetVal 326
 Db 1361 ATAGACCTTAACAGGAGAGAGATAACAGGTATCTGAACCACTGGGTCTACCATGATA 1420
 QY 327 GluHisAsnTrpTrpValSerAlaTrpArgIleLeuTrpSerAspAspGlyGlnLysTrp 346
 Db 1421 GAACACAGTTACTATGTCTGCTACAGAGTCTGTACAGTGCAGATGGGCAGAGATGG 1480
 QY 347 ThrValTrpArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTrp 366
 Db 1481 ACTGTGTACAGAACTGGTGTGGACAGGACAAAGATATTTCAAGGAAACAAAGATTAT 1540
 QY 367 HisGlnAspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgValAsn 386
 Db 1541 CACAAGGATGTCTGAATAACTTTTGGCCCACTATTATGCACGTTTCATTAGAGTGAAC 1600
 QY 387 ProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheIle 406
 Db 1601 CCTGTCCAGTGGCAACAGAAATGGCATGAAAGTGGAGTCTCGGATGTCAGTTACT 1660
 QY 407 ProLysGlyArgProProLysLeuThrGlnProProProProArgAsnSerAsnAspLeu 426
 Db 1661 CTCAAAGGTGGCTTCCTCAAGCTTACT-----CCACCTCTCGGAAACGGCAATAACCTC 1714
 QY 427 LysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPheThrGlnPro 446
 Db 1715 AGAATTAATACAGCTCTCCCAACTAGTAAAGTGTGCTCCCTCAATTTACTCAAGTG 1774
 QY 447 LeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerPro 466
 Db 1775 CTCCAACCTCGAAGTGAATGAATCTCTGTGCGAGCGGGGAGACAACTACCACCTCT 1834
 QY 467 AspIleArgAsnThrThrValThrProAsnValThrLysAspValAlaLeuAlaVal 486
 Db 1835 GATATAAAACACGACTGTAACTCCAGTGTAAACCAAGATGTCGCACCTGGCTGCCGTT 1894
 QY 487 LeuValProValLeuValMetValLeuThrThrLeuIleLeuIleLeuValLysAlaTrp 506
 Db 1895 CTGTCTCCTGTGCTGTCATGGCCCTCACCACACTATCCTCATCTAGTGTGCTGG 1954
 QY 507 HisTrpArgAsnArgLysLysThrGluGlyThrTrpAspLeuProTrpTrpAspArg 526
 Db 1955 CACTGGAGAACAGGAGAAACCTGAAGCGCTATGATTTACCCCACTGGGATCGG 2014
 QY 527 AlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGlu 546
 Db 2015 GCAGGTGGTGGGAAGGAATGAAGCAGTCTTCCCTGCGCAAGTGGGTGGACGAGGAG 2074
 QY 547 ThrProValArgTrpSerSerGluValAsnHisLeuSerProArgGluValThrThr 566
 Db 2075 ACGCCAGTCCCTACAGCACTAGTGAAGTCACTGAGTGGCGGAGAGTACCACCA 2134
 QY 567 ValLeuGlnAlaAspSerAlaGluTrpAlaGlnProLeuValGlyGlyIleValGlyThr 586
 Db 2135 GTGTGTCAGCGGACTCTGCAGATATATGCACAGCCCTCGTGGGAGGAATGTTGGCACA 2194
 QY 587 LeuHisGlnArgSerThrPheLysProGluGlyLysGluAlaGlyTrpAlaAspLeu 606

Db 2195 CTCATCAGAGATCCACCTTTAAACCTGAGGAAGGAGGAGGCTATGCAGACCTC 2254
 QY 607 AspProTrpAsnSerProGlyGlnGluValTrpHisAlaTrpAlaGluProLeuProIle 626
 Db 2255 GATCCTTACAACTCTCCAAATGCAGGAAGTGTACACGCTATGCTGNAACCACTGCCGTA 2314
 QY 627 ThrGlyProGluTrpAlaThrProIleIleMetAspMetSerGlyHisProThrThrSer 646
 Db 2315 ACGGGGCTCAGTACGCAACCCCGATGTCATGCATGTCCAGGACCCACACAGCCTCA 2374
 QY 647 ValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProProProLeuVal 666
 Db 2375 GTTGGTCTGCCCTCCACATCCACTCCAAACTCCAGGACCCAGCCCTCAGCTTTAGTG 2434
 QY 667 GlyThrTrpAsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTrp 686
 Db 2435 GGAATTAACAACACTCTTCTCTCCAGGACTGACAGCTGTTCTCCAGCCAGGCTCAGTAT 2494
 QY 687 AspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTrpGlnVal 706
 Db 2495 GACACCCCAAAAGGTGGGAAG---TCAGCTCTACCCAGGAGGAACCTGTATACCAAGTG 2551
 QY 707 ProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGlyGluCysAspValPheLys 726
 Db 2552 CCCAGAGACCCAGGAGCTATCAGGAGCAGGAGGATGAGAGTGTGCTTTTAA 2611
 QY 727 GluIleLeu 729
 Db 2612 GAAATCCTT 2620
 RESULT 6
 AK056350
 LOCUS
 DEFINITION
 Homo sapiens cDNA FLJ1788 fls. clone NT2R12008598, weakly similar to MILK FAT GLOBULE-EGF FACTOR 8 PRECURSOR.
 ACCESSION
 AK056350.1 GI:16551729
 VERSION
 AK056350.1
 KEYWORDS
 Homo sapiens teratocarcinoma cell line:NT2 cDNA to mRNA,
 clone.lib:NT2R12 clone:NT2R12008598.
 SOURCE
 Homo sapiens
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Theria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1
 Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S., Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S., Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M., Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuhio,Y., Nagai,K. and Isogai,T.
 NEDO human cDNA sequencing project
 TITLE
 UNpublished
 JOURNAL
 2 (bases 1 to 2020)
 REFERENCE
 Isogai,T., Otsuki,T. and Sugiyama,T.
 AUTHORS
 Direct Submission
 TITLE
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 JOURNAL
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.
 FEATURES
 Location/Qualifiers
 1..2020
 /organism="Homo sapiens"

QY 205 GlyGlyGlnIleSerValIleSerLysGlyIleProTyrTyrGluSerSerLeuAla 224
|||||
Db 640 GGTGCCAGATCATGCTGCCTCAGCGCAAGGGATCAGTCATATGAGGGATCTGGCC 699
QY 225 AsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThrPheLysThrSer 244
|||||
Db 700 AATGCTGTTCTTCCGGGATGCTCCCTGTCAGACAGCGATCTCTGTTTACCTCCAA 759
QY 245 GlyCysTyrGlyThrLeuGlyMetGluSerClyValIleAlaAspProGlnIleThrAla 264
|||||
Db 760 GGTTCACAGATCCTTACGTTTGA-----CTGACGGGCAATCAGAGCT 807
QY 265 SerSerValLeuGluTyrThrAspHisThrGlyGlnGluAsnSerTyrLysProLysLys 284
|||||
Db 808 TCTTCTCATGGCAGTCGGTCAATGAGAGTGGAGACCAAGTTCAGTGTCTCTGGCCAA 867
QY 285 AlaArgLeuLysLysProGlyProProTrpAlaAlaPheAlaThrAspGlu----- 301
|||||
Db 868 GCCCGACTTCAGGACCAAGGCCCATCATGCGCTTCGGCGGACAGTAGCAACAACACAAA 927
QY 302 --TyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThr 320
|||||
Db 928 CCAGCAGAGTGGCTGGAGATCGATTGGGGAGAAAGAAATAACAGGAATTAGGACC 987
QY 321 ThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSer 340
|||||
Db 988 ACAGATCTACAGATCGAATCTCACTTTATGTTAGAGTTTGTGATGAATCTCAA 1047
QY 341 AspAspGlyGlnLysTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePhe 360
|||||
Db 1048 AACATAATTCTAAGTGAAGACCTATAAAGGAATTGTGAATAATAAGAAAGGTGTT 1107
QY 361 GlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProProIleIleAla 380
|||||
Db 1108 CAGGTAACCTAATCTTCGGGACCCAGTCGCAACAAATTTCACTCCCTCCATCGTGGCC 1167
QY 381 ArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeu 400
|||||
Db 1168 AGATATGCGGGTGTCCCCCAACATGGCCAGCAGGATAGCTTGAAGTGGAGCTC 1227
QY 401 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProPro 420
|||||
Db 1228 ATTGTTGCCAG-----ATTACAAA----- 1248
QY 421 ArgAsnSerAsnAsp-----LeuLysAsnThrThrAlaProProLysIleAla 436
|||||
Db 1249 -----GGTAATGATTCATGTTGGCGCAAGACAAGTCAAGCACCAGTGTTCAACT 1302
QY 437 LysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPhePro 456
|||||
Db 1303 AAGAAAGAAGATGAGACAATCACAAGGCCCATC----- 1335
QY 457 AlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsn 476
|||||
Db 1336 ---CCCTCGGAAGAAACA-----TCCACAGGAATAAACAATTACAACG----- 1374
QY 477 ValThrLysAspValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThr 496
|||||
Db 1375 -----GTGGTATTCCATGCTGCTCTCTGTTGTCCTGGTGTTCCTGGGAATG 1422
QY 497 ThrLeuIleLeuIleLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGlu 516
|||||
Db 1423 GGGATCTTCAGCCTTT-----AGAAAGAAAGAAAGAAA 1458
QY 517 GlyThrTyrAspLeuProTyrTrp-----AspArgAlaGlyTrpTrpLysGly 532
|||||
Db 1459 GGAAGT-----CCGTATGGATCAGCGGAGCTCAGAAACACAGACTGTGGAAGCAG 1509
QY 533 MetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSer 552
|||||
Db 1510 ATTAATATATCCCTTT-----GCCAGACATCAG 1536

QY 553 SerSerGlu-----ValAsnHisLeuSerProArgGluValThrThrValLeuGlnAla 570
|||||
Db 1537 TCACCTGAGTTTACCATCAGCTATGATAATGAGAGGAGATGACACAAAGTTAGATCTC 1596
QY 571 AspSerAlaGluTyrAla 576
|||||
Db 1597 ATCACAAGTGATGGCA 1614
RESULT 11
AX118818
LOCUS AX118818 1761 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 3 from Patent WO0129219.
ACCESSION AX118818
VERSION AX118818.1 GI:14035775
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1761)
AUTHORS Donoho G., Turner, C.A., Nehls, M., Friedrich, G., Zambrowicz, B. and
Sands, A.T.
TITLE Human cub-domain-containing protein and gene encoding the same
JOURNAL Patent: WO 0129219-A 3 26-APR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source
1..1761
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 471 a 402 c 483 g 405 t
ORIGIN
Alignment Scores:
Pred. No.: 1.03e-54 Length: 1761
Score: 924.50 Matches: 220
Percent Similarity: 54.27% Conservative: 98
Best Local Similarity: 37.54% Mismatches: 187
Query Match: 23.78% Indels: 81
DB: 6 Gaps: 16
US-10-060-830-3 (1-729) x AX118818 (1-1761)
QY 5 LeuLeuLeuLeuLeuValLeuValLeuLeuLeuLeuGluAspAlaGlyAlaGlnGlnGlyAsp 24
|||||
Db 199 CTGCTTGTGCTGCTGCGGCTCCGCCCGCTCCGCGTGCAGCGGAGGAGCTGGGTGAT 258
QY 25 GlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerIleAsnTyrPro 44
|||||
Db 259 GGCCTGTGCACACCTAGTACTTATCAGATAGTGGCACAATGACATCTAAGAATTATCCC 318
QY 45 GlnThrTyrProAsnSerThrValCysGluTrpGluIleArgValLysMetGlyGluArg 64
|||||
Db 319 GGGACCTACCCCAATCACACTGTTTGGAAAGACAATACAGTACCAAGGGGAAAGA 378
QY 65 ValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCysHisPheAsnTyr 84
|||||
Db 379 CTGATTCGAGGTTGGGAGATTTCGATATCGAA---TCCCAGACCTGTGCTTCGACTAT 435
QY 85 LeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLysTyrCysGlyLeu 104
|||||
Db 436 CTTCTCTTCCACAGC-----TCITCAGATCAATATGTTCCATCTGTGGA--- 480
QY 105 GlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPheMet 124
|||||
Db 481 AGTATGACTGTTCACCAAGAACTCTTGTGTAACAACAAGTGAAGTAACCGTCCGCTTGAG 540
QY 125 SerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGln 144
|||||
Db 541 AGTGGATCCCAATTTCTGCCCGGGGTTTTTGTGCTACCTATGCGACGAGGACCATCCA 600
QY 145 AspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyr 164
|||||
Db 601 GATTTAATACATGTTTGGAAACGAGCTAGCCATTATTGTAAGACAGAAATACAGCAAAATTC 660

QY 165 CysProAlaGlyCysLeuLeuProPheAlaGluLeuSerGlyThrIleProHisGlyTyr 184
 Db 661 TGCCAGCTGTTGTAGAGAGCTAGCAGGACATTTTCTGGTAATATGGTAGATGGATAT 720
 QY 185 ArgAspSerProLeuCysMetAlaGlyValHisAlaGlyValValSerAsnThrLeu 204
 Db 721 AGAGATACCTCTTTATGTGCAAGCTGCCATCCATCCAGGAATAATTGCTGATGAACATA 780
 QY 205 GlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGluSerSerLeuAla 224
 Db 781 GGTGCCAGACAGTCTGCTTCAGCGCAAGGATCATGATATGAAGGATTCGTGCC 840
 QY 225 AsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThrPheLysThrSer 244
 Db 841 AATGGTGTCTTCGAGGAGTGTCTCCCTGTGCAGACAAGCATTTCTGTTTACCTCCAAT 900
 QY 245 GlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAla 264
 Db 901 GGTTCGACAGATCTCTGAGTTTGA-----CTGACGGGCAAAATCAGAGCT 948
 QY 265 SerSerValLeuGluThrThrAspHisThrGlyGlnGluAsnSerTyrLysProLysLys 284
 Db 949 TCTTCTCATGCGATCGTCAATGAGTGGAGACCAAGTCTACTGCTCTCTGGCCAA 1008
 QY 285 AlaArgLeuLysProGlyProProThrAlaAlaPheAlaThrAspGlu----- 301
 Db 1009 GCCCGACTTCAGGACCAAGGCCCATCATGGCTTCGGCGCAGATGAGCAACACACAAA 1068
 QY 302 ---TyrGlnThrIleGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThr 320
 Db 1069 CCACGAGAGTGTGGAGATCATTTGGGGGAGAAAAAGAAATAACAGGAATTAGGACC 1128
 QY 321 ThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSer 340
 Db 1129 ACAGGATCATACAGCTCGAATCTCACTTTATGTTAAGAGTTTGTGATGAATCTCAA 1188
 QY 341 AspAspGlyGlnLysThrValTyrArgGluProGlyValGluGlnAspLysIlePhe 360
 Db 1189 AACAATAATCTAAGTGAAGACCTATAAAGGAATTGTAATAATCAAGAAAGTGTTT 1248
 QY 361 GlnGlyAsnLysAspThrHisGlnAspValArgAsnAsnPheLeuProIleIleAla 380
 Db 1249 CAGGGTAATCTAACTTTCGGACCCAGTGCACAAACAATTTATCCCTCCCATCGTGCC 1308
 QY 381 ArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeu 400
 Db 1309 AGATATGTCGGGTGTCCTCCAGACATGCGCAGAGGATAGCCTTGAAGTGGAGCTC 1368
 QY 401 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProPro 420
 Db 1369 ATGGTGTGCAG-----ATTACAAA----- 1389
 QY 421 ArgAsnSerAsnAsp-----LeuLysAsnThrThrAlaProProLysIleAla 436
 Db 1390 -----GGTAATGATTTCATGTGTGGCGCAAGCAAGTCAAGACACACAGTGTTCACAT 1443
 QY 437 LysGlyArgAlaProLysPheThrGlnProLysGlnProArgSerSerAsnGluPhePro 456
 Db 1444 AAGAAGAAGATGAGACATCAACAGGCCATC----- 1476
 QY 457 AlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsn 476
 Db 1477 ---CCCTCGGAAGAAACA-----TCCACAGGAATAAATACATTAACAG----- 1515
 QY 477 ValThrLysAspValAlaLeuAlaValLeuValProValLeuValMetValLeuThr 496
 Db 1516 -----GTGGCTATTCCATGGTGTCTCTGTTGTCTGCTGTGTCTGCTGGAATG 1563
 QY 497 ThrLeuIleLeuIleLeuValCysAlaThrIleHisThrArgAsnArgLysLysThrGlu 516
 Db 1564 GGGATCTTTCAGCCTTT-----AGAAAGAGAAGAAAGAAA 1599

QY 517 GlyThrTyrAspLeuProTyrTrp-----AspArgAlaGlyTyrTrpLysGly 532
 Db 1600 GGAAGT-----CCGTATGGATCAGCGAGGCTCAGAAAACAGACTGTTGAAGCAG 1650
 QY 533 MetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSer 552
 Db 1651 ATTAATATCCCTTT-----GCCAGACATCAG 1677
 QY 553 SerSerGlu-----ValAsnHisLeuSerProArgGluValThrThrValLeuGlnAla 570
 Db 1678 TCAGCTGAGTTTACCATCAGCTATGATAATGAGAGGAGATGACACAAAGTTAGATCTC 1737
 QY 571 AspSerAlaGluTyrAla 576
 Db 1738 ATCAAGATGATATGGCA 1755
 RESULT 12
 AC013497 192247 bp DNA linear HTG 26-MAR-2000
 LOCUS Homo sapiens chromosome 3 clone RP11-319J24 map 3, WORKING DRAFT
 DEFINITION Homo sapiens chromosome 3, clone RP11-319J24
 AC013497
 VERSION AC013497.4 GI:7329299
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 192247)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 3, clone RP11-319J24
 Unpublished
 2 (bases 1 to 192247)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckhgalter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collins, S., Collins, S.,
 Cook, P., Dearlano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
 Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hages, B., Hesford, A., Horton, L.,
 Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehocsky, J., Liu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meidrin, J.,
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 26, 2000 this sequence version replaced g1:6692347.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1852
 Center clone name: 319_J24
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 182870 bases at least Q40
 Consensus quality: 187032 bases at least Q30
 Consensus quality: 188925 bases at least Q20
 Insert size: 194000; agarose-1p
 Insert size: 191247; sum-of-contigs
 Quality coverage: 5.0 in Q20 bases; agarose-1p
 Quality coverage: 5.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 472: contig of 472 bp in length
* 473 572: gap of 100 bp
* 573 3945: contig of 3373 bp in length
* 3946 4045: gap of 100 bp
* 4046 11056: contig of 7011 bp in length
* 11057 11156: gap of 100 bp in length
* 11157 17728: contig of 6572 bp in length
* 17729 17828: gap of 100 bp
* 17829 30556: contig of 12728 bp in length
* 30557 30656: gap of 100 bp
* 30657 49294: contig of 18638 bp in length
* 49295 49394: gap of 100 bp
* 49395 64509: contig of 15115 bp in length
* 64510 64609: gap of 100 bp
* 64610 83042: contig of 18433 bp in length
* 83043 83142: gap of 100 bp
* 83143 112373: contig of 29231 bp in length
* 112374 112473: gap of 100 bp
* 112474 145663: contig of 33190 bp in length
* 145664 145763: gap of 100 bp
* 145764 192247: contig of 46484 bp in length.

FEATURES

source

1. 192247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-319J24"
/clone_lib="RP11 Human Male BAC"

misc_feature 1. 472

/note="assembly_fragment"
clone_end:SP6
vector_side:right"

misc_feature 573..3945

/note="assembly_fragment"

misc_feature 4046..11056

/note="assembly_fragment"

misc_feature 11157..17728

/note="assembly_fragment"

misc_feature 17829..30556

/note="assembly_fragment"

misc_feature 30657..49294

/note="assembly_fragment"

misc_feature 49395..64509

/note="assembly_fragment"

misc_feature 64610..83042

/note="assembly_fragment"
clone_end:T7
vector_side:left"

misc_feature 83143..112373

/note="assembly_fragment"

misc_feature 112474..145663

/note="assembly_fragment"

misc_feature 145764..192247

/note="assembly_fragment"

BASE COUNT 58524 a 38157 c 36840 g 57721 t 1005 others
ORIGIN

Alignment Scores:

Pred. No.: 4-87e-52 Length: 192247
Score: 922.50 Matches: 201
Percent Similarity: 44.87% Conservative: 0
Best Local Similarity: 44.87% Mismatches: 1
Query Match: 23.73% Indels: 247

DB: 2 Gaps: 1
US-10-060-830-3 (1-729) x AC013497 (1-192247)
QY 528 GlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThr 547
|||||
Db 164102 GGTGGTGGAAAGGAATGAAGCAGTTCTCTCCCTGCAAAAGCAGTGGACCATGAGGAACC 164161
QY 548 ProValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrVal 567
|||||
Db 164162 CCAGTTCGGCTATACGACGACGGAAGTTAATCACCTGAGTCCAAAGAGAAGTCCACACAGT 164221
QY 568 LeuGlnAlaAspSerAla----- 573
Db 164222 CTGCAGGCTGACTCTGC-AGGTAACATATGTTGCAGCCTTCTGGTACCAGCAGAGGAGA 164280
QY 573 ----- 573
Db 164281 AACTGCTTAGGCTTGCTATAAAGTGCTTTGGGATTTACAGTTCTTTGATCCCTTTCATGT 164340
QY 573 ----- 573
Db 164341 TTAAGAAATGAATGTTTCTCCTAAAGGTAGAACCACTTTTAAAGGTGACTCTTCAGAG 164400
QY 573 ----- 573
Db 164401 AGTCATTTCTGTATCTTGGATATTCGTCTGTAAGTTGGGACATGATTCTTAAAGACT 164460
QY 573 ----- 573
Db 164461 CTGGCATTATGATGTTAACTGAGGCTTGTGGGAGACCCGAGATGTTGAGAGGAGAGC 164520
QY 573 ----- 573
Db 164521 GGGAAAGACAGAAAGACCTTACAGTGTTCCTCTTACGCTTAGTACATACCTATTTTA 164580
QY 573 ----- 573
Db 164581 AACATTTCCTAATATTGAAAAAAATCTCTGAAATTAATTTTCAAGATTAATATTATTC 164640
QY 573 ----- 573
Db 164641 CAGTAACGTAATGTTCTCTTTTCCAATGAAATATTCGAAATTAATTCAGAAAGGAATAT 164700
QY 573 ----- 573
Db 164701 TCAGTAAAGAAATATGTTGCTGTAATCTAAAGACATATGTGAAAAATTCATATATGCAA 164760
QY 573 ----- 573
Db 164761 AATTACATGAATCAGACTTCTCTGGTATATTATTACTATGAGAAGTACTAGACAGTATGTT 164820
QY 573 ----- 573
Db 164821 TAATTTTAATTTCAAAGTCTTTACAGATTAAAGTATTGATGGCAACAAAAGTATCCTAAA 164880
QY 573 ----- 573
Db 164881 GAATGTTTGGCTTTTAGTATTAATCTGCAGCTGCCACCCCTAAATATGCAATCTCTTACTA 164940
QY 574 ----- 574
Db 164941 GTAAATGACTTTTCTCTGTTTGTGATATCCACAGAGTATCTCTCAGCCACTGGTAGGA 165000
QY 582 GlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGluAla 601
|||||
Db 165001 GGAATTTGGTACACTTCAATCAAGATCTACCTTTAAACACAGAGAAGGAAAGAAGCA 165060
QY 602 GlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyrAla 621
|||||
Db 165061 GGCTATGCAGACGATAGTATCCTTACAACCTCACCAGGCGAGAGTTATCATGCTATGCT 165120
QY 622 GluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetaspMetSerGly 641


```
repeat_region complement(4203..4581)
repeat_region /rpt_family="L2"
repeat_region 4767..4790
repeat_region /rpt_family="(TC)n"
repeat_region 4791..4851
repeat_region /rpt_family="(TC)n"
repeat_region 4854..4892
repeat_region /rpt_family="(CA)n"
repeat_region 4907..5066
repeat_region /rpt_family="L2"
repeat_region 5333..5386
repeat_region /rpt_family="GA-rich"
repeat_region complement(5512..5945)
repeat_region /rpt_family="MLTIC"
repeat_region complement(6187..6624)
repeat_region /rpt_family="LIR37B"
repeat_region complement(6858..7219)
STS 6880..7000
repeat_region /standard_name="73424"
repeat_region complement(7409..7632)
repeat_region /rpt_family="MIR"
repeat_region 7822..8369
repeat_region /rpt_family="L1MC1"
repeat_region 8417..8524
repeat_region /rpt_family="L1M4"
repeat_region complement(10026..10215)
repeat_region /rpt_family="MER20"
repeat_region complement(10485..10549)
repeat_region /rpt_family="MER112"
repeat_region 12814..13590
repeat_region /rpt_family="L1PA7"
repeat_region 13917..14135
repeat_region /rpt_family="L1MA5A"
repeat_region complement(14753..15070)
repeat_region /rpt_family="AluSq"
repeat_region 15152..15366
repeat_region /rpt_family="MIR"
repeat_region 15395..15562
repeat_region /rpt_family="MER5B"
repeat_region 16109..16388
repeat_region /rpt_family="L1MC5"
repeat_region 16389..16642
repeat_region /rpt_family="AluY"
repeat_region 16643..16758
repeat_region /rpt_family="L1MC5"
repeat_region 16826..16898
repeat_region /rpt_family="L1MC5"
repeat_region complement(17591..18655)
repeat_region /rpt_family="L1MC4"
repeat_region complement(19191..19653)
repeat_region /rpt_family="L1MD3"
repeat_region complement(19662..19780)
repeat_region /rpt_family="FLAM_A"
repeat_region complement(19781..21049)
repeat_region /rpt_family="L1MD3"
repeat_region complement(21069..21455)
repeat_region /rpt_family="MLTIB"
repeat_region complement(21479..22318)
repeat_region /rpt_family="L1MD3"
repeat_region complement(22447..22681)
repeat_region /rpt_family="L1MD3"
repeat_region complement(23470..23853)
repeat_region /rpt_family="MSTD"
repeat_region 23852..23925
repeat_region /rpt_family="U3"
repeat_region 23936..25203
STS 24844..24975
repeat_region /standard_name="6234"
repeat_region 25219..25388
repeat_region /rpt_family="(TTTC)n"
repeat_region complement(25428..25548)
```

```
repeat_region /rpt_family="AluJb"
repeat_region 25555..25950
repeat_region /rpt_family="L1PB3"
repeat_region 25951..26456
repeat_region /rpt_family="MER75"
repeat_region 26457..27565

Alignment Scores:
Pred. No.: 4.91e-52 Length: 193623
Score: 922.50 Matches: 201
Percent Similarity: 44.87% Conservative: 0
Best Local Similarity: 44.87% Mismatches: 1
Query Match: 23.73% Indels: 247
DB: 9 Gaps: 1

US-10-060-830-3 (1-729) x AC106728 (1-193623)
Qy 528 GlyTrrPlyGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThr 547
Db 95315 GGTGGTGGAAAGGAATGAGCAGTTTCTCTGCAAAAGCAGTGGACCATGAGGAACC 95374
Qy 548 ProValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrVal 567
Db 95375 CCAGTTCGCTATAGCAGCAGCGAGATTAACTACCTGAGTCCAGAGAGAAAGTCCACACAGTG 95434
Qy 568 LeuGlnAlaAspSerAla----- 573
Db 95435 CTGCAGGCTGACTTCGC-AGGTAAGTATGTTGCAGCCTTCTGGTACCAGGAGAGGGAGA 95493
Qy 573 ----- 573
Db 95494 AACTGCTTAGGCTTGCTATAAGTCTTTGGGATTACAGTTCCTTTGATCCCTTCATGT 95553
Qy 573 ----- 573
Db 95554 TTAAGAAATGAATTGTTTCTTAAAGGTAGAACCACTTTTTTAAAGGTGACTCTTCAGAG 95613
Qy 573 ----- 573
Db 95614 AGTCATTTCTGTATCTTGGATATTCGTCTGTAAGTTGGGACATGATTCTTAAAGACT 95673
Qy 573 ----- 573
Db 95674 CTGGCATTATGATGTTAACTGAGGCTTGTTGGGAGACCCGAGATGTTGAGAGGAGAGC 95733
Qy 573 ----- 573
Db 95734 GGGAAAGACAGAAGAAGCCTTACAGTGTTTATCCTCTTAGCCTAGTACATACTATTATTTA 95793
Qy 573 ----- 573
Db 95794 AACATTTCTTAATATTTGAAAAAATCTCTGAAATATTTTCAAGATTAATATTTATTC 95853
Qy 573 ----- 573
Db 95854 CAGTAAGTGAATGTTTCTTTTCCAAATGAATATTTCTGAAATAATTCAGAAAGGAATAT 95913
Qy 573 ----- 573
Db 95914 TCAGTAAGAAATATGTTGCTGTAAATCTTAAAGAACATATGTGAAATTTCTATATGCAA 95973
Qy 573 ----- 573
Db 95974 AATTACATGAATGAGACTTTCTGGTATATATTACTATGAGAAGTACTAGACAGTATGTT 96033
Qy 573 ----- 573
Db 96034 TAATTTTAATTTTCAAAGTCTTACAGATTAAAGTATTGATGGCAACAAAAGTATCTTAAA 96093
Qy 573 ----- 573
Db 96094 GAATGTTTGGCTTTTAGTATATAATCTGCACCTGCCACCCCTAAATATTATGCAATCTCTTTACTA 96153
```

```
QY 574 -----GluTyrAlaGlnProLeuValGly 581
Db 96154 GTAAATGACCTTTCTGTTGTTTCTGATATCCACAGAGTATGCTCAGCCACTGGTAGGA 96213
QY 582 GlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGluAla 601
Db 96214 GGAATTTGTTGGTACACTTCATCAAGATCTACCTTTAAACCAAGAAGAAGAAAGCA 96273
QY 602 GlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyrAla 621
Db 96274 GGCTATGACAGCTAGATCCTTACAACTCACCAGGCGAGGAAGTTTATCATGCCCTATGCT 96333
QY 622 GluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetAspMetSerGly 641
Db 96334 GAACCACTCCCAATACGGGGCTCAGTATGATCAACCCCAATCATCATGACATGTCAGG 96393
QY 642 HisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGln 661
Db 96394 CACCCACAACTCAGTTGGTCAGCCCTCCACATCCACTTTCAGGCTACGGGGAACCAA 96453
QY 662 ProProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSer 681
Db 96454 CTTCCCCCACTAGTGGAACTTACATACATACATCTCTCCAGGACGTGACAGCTGCTCCCA 96513
QY 682 AlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAspGlu 701
Db 96514 GCCCAGGCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGCTCCTACCTGCCCCAGACGAA 96573
QY 702 LeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGlyGlu 721
Db 96574 TTGGTACAGGTGTCACAGACACAAAGAAGTATCAGGAGCAGGAAGGATGGGAA 96633
QY 722 CysAspValPheLysGluIleLeu 729
Db 96634 TGTGATGTTTTTAAAGAAATCCTT 96657

RESULT 14
AX118822
LOCUS AX118822 1768 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 7 from Patent WO0129219.
ACCESSION AX118822
VERSION AX118822.1 GI:14035777
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Donoho, G., Turner, C.A., Nehls, M., Friedrich, G., Zambrowicz, B. and
Sands, A.T.
TITLE Human cub-domain-containing protein and gene encoding the same
JOURNAL Patent: WO 0129219-A 7 26-APR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source
location/Qualifiers
BASE COUNT 539 a 370 c 421 g 438 t
ORIGIN
Alignment Scores:
Pred. No.: 3,68e-54 Length: 1768
Score: 916.50 Matches: 215
Percent Similarity: 54.75% Conservative: 96
Best Local Similarity: 37.85% Mismatches: 176
Query Match: 23.57% Indels: 81
DB: 6 Gaps: 16

US-10-060-830-3 (1-729) x AX118822 (1-1768)
QY 23 GlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerIleAsn 42
Db 14 GGTGATGGCTGTGGACACCTAGTACTTATCAGGATAGTGGCACAATGACATCTAAGAAT 73
```



```
QY 394 IleAlaMetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLys 413
Db 1051 ATAGCCTTGAAGTGGAGCTCATTTGTTGCCAG----- 1083
QY 414 LeuThrGlnProProArgAsnSerAsnAsp-----LeuLysAsnThr 429
Db 1084 ATTACACAA-----GGTAATGATTCATTTGTTGGCGCAAGACAAGT 1125
QY 430 ThrAlaProProLysIleAlaLysGlyArgAlaProLysPheThrGlnProLysGlnPro 449
Db 1126 CAAAGCACCAGTGTTCACCTAAGAAGAAGATGAGACAATCACAGGCCCATC----- 1179
QY 450 ArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerProAspIleArg 469
Db 1180 -----CCCTCGGAAGAACA-----TCCACAGGAATAAAC 1209
QY 470 AsnThrThrValThrProAsnValThrLysAspValAlaLeuAlaValLeuValPro 489
Db 1210 ATTACACG-----GTGGCTATCCCATTTGTTCCCTTGTT 1245
QY 490 ValLeuValMetValLeuThrThrLeuIleLeuIleLeuValCysAlaTrpHisTrpArg 509
Db 1246 GTCCTGGTGTTCCTGGAATGGGATCTTTGCAGCCTTT----- 1284
QY 510 AsnArgLysLysThrGluGlyThrTyThrAspLeuProTyTrp-----Asp 525
Db 1285 ---AGAAAGAAGAAGAAGAAAGAGT-----CCGTATGGATCAGCGAGGCTCAG 1332
QY 526 ArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGlu 545
Db 1333 AAAACAGACTGTTGGAAGCAGATTAAATATCCCTTT----- 1368
QY 546 GluThrProValArgTyrSerSerGlu-----ValAsnHisLeuSerProArgGlu 563
Db 1369 -----GCCAGACATCAGTCAGCTGAGTTTACCATCAGCTATGATATGAGAAGGAG 1419
QY 564 ValThrThrValLeuGlnAlaAspSerAlaGluTyrAla 576
Db 1420 ATGACACAAAAGTTAGATCTCATCACAGTGATATGGCA 1458
```

Search completed: January 21, 2003, 12:55:50
Job time : 5227.02 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 23, 2003, 12:48:26 ; Search time 2592.12 Seconds
(without alignments)
4554.775 Million cell updates/sec

Title: US-10-060-830-3
Perfect score: 3888
Sequence: 1 MPFLFLLLVLLLLLEDAGA.....TQEVSGAGRGCDVDFKEIL 729

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US10060830/runat_23012003_124821_6284/app_query.fasta_1.1742
-DB=EST -OPWT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830@cgn.1.1.3526 -runat_23012003_124821_6284 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-FGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_estl: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1494.5	38.4	1065	13	BM450083
2	1327.5	34.1	1061	14	BM905706
3	1236	31.8	694	13	BI860608
4	1212	31.2	712	9	AI378788
5	1151.5	29.6	1272	11	AK006805
6	1090.5	28.0	641	9	AI435602
7	1087	28.0	680	14	BQ014739
8	1072	27.6	693	12	BF732384
9	1067	27.4	608	12	BE178371
10	1064	27.4	618	12	BE894226
11	1047.5	26.9	671	9	AA147037
12	1037	26.7	643	10	BB179946
13	1019	26.2	703	10	AW070902
14	1018	26.2	692	12	BF732861
15	1011	26.0	664	10	AW383889
16	990	25.5	600	14	BM781107
17	977	25.1	608	10	BB657747
18	974	25.1	575	10	AV603144
19	974	25.1	912	12	BF167024
20	972	25.0	628	14	BQ037529
21	956	24.6	560	12	BF074573
22	940	24.2	535	9	AI417256
23	932	24.0	605	10	AW138866
24	915	23.5	667	9	AL586285
25	898	23.1	575	10	AW383902
26	880	22.6	587	10	AW383890
27	877	22.6	538	14	N21309
28	865	22.2	566	12	BF34973
29	850	21.9	566	9	AA147128
30	835	21.5	597	9	AJ450411
31	808	20.8	526	9	AJ420312
32	802	20.6	459	9	AA885289
33	791	20.3	457	9	AI276772
34	790	20.3	466	9	AI750960
35	780	20.1	525	9	AI290500
36	778.5	20.0	689	9	AJ444823
37	778.5	20.0	764	9	AJ441638
38	778	20.0	459	14	N46066
39	776	20.0	504	9	AI565996
40	772.5	19.9	809	9	AJ445175
41	770	19.8	463	12	BF777080
42	757	19.5	502	9	AI499115
43	756	19.4	574	13	BJ032248
44	753	19.4	778	9	AJ441380
45	750.5	19.3	817	9	AJ396784

ALIGNMENTS

RESULT 1
BM450083

LOCUS BM450083 1065 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6393434 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528102
5', mRNA sequence.

ACCESSION BM450083
VERSION BM450083.1 GI:18499123
KEYWORDS EST.

SOURCE Human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1065)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCFD/DTF
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12204 row: g column: 15
High quality sequence stop: 665.
Location/Qualifiers
1..1065
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5528102"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SpOrt6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."

FEATURES

source

1..1065

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5528102"

/clone_lib="NIH_MGC_72"

/tissue_type="melanotic melanoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pCMV-SpOrt6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 312 a 280 c 236 g 237 t

ORIGIN

Alignment Scores:

Pred. No.: 8,44e-140 Length: 1065
Score: 1494.50 Matches: 284
Percent Similarity: 98.28% Conservative: 1
Best Local Similarity: 97.93% Mismatches: 4
Query Match: 38.44% Indels: 1
DB: 13 Gaps: 1

US-10-060-830-3 (1-729) x BM450083 (1-1065)

QY 439 ArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGln 458
DB 1 CGTGCCCCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATTTCTCTGCACAG 60
QY 459 ThrGluGlnThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThr 478
DB 61 ACAGACAACAACTGCCAGTCTGTATATACAGATACTACCTACATCCCAATGTAACT 120
QY 479 LysAspValAlaLeuAlaValLeuValProValLeuValMetValLeuThrLeu 498
DB 121 AAAGATGTAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 499 IleLeuIleLeuValCysAlaIleThrPheArgAsnArgLysLysLysThrGluGlyThr 518
DB 181 ATTCTCATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 519 TyrAspLeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuPro 538
DB 241 TAGACTTACCTTACTGGACCGGGCAGGTGGTGGAAAGGATGAGCAGTTCTCTCT 300
QY 539 AlaLysAlaValAspHisGluGluThrProValArgTyrSerSerSerGluValAsnHis 558
DB 301 GCAAAAGCAGTGGACCATGAGGAACCCAGTTCGCTATAGCAGCAGCGAAGTAAATCAC 360
QY 559 LeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGluThrAlaGlnPro 578
DB 361 CTGAGTCCAAAGAGAACTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 579 LeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGly 598
DB 421 CTGCTAGGAGGAATGTTGGTACACTTCATCAAGATCTACCTTTAAACCAAGAAGA 480
QY 599 LysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHis 618
DB 481 AAAGAAGCAGGCTATGCAGACCTAGATCCTTACAACTCACCAGGGCAGGAAGTTATCAT 540

QY 619 AlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetasp 638
DB 541 GCCATATGCTGAACCACTCCCAATTTACGGGGCTAGTATGCAACCCCAATCATCATGAC 600
QY 639 MetSerGlyHisProThrThrSerValGlyGlnProSerThrThrPheLysAlaThr 658
DB 601 ATGTGAGGACCCCAACATTCAGTTGGTCAGCCCTCCACATCCACITTTCAAGGCTAGC 660
QY 659 GlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSer 678
DB 661 GGAACCAACCTCCCCACTAGTGGAACTTACAAATACACTTCTCTCCAGGACTGACAGC 720
QY 679 CysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuProAla 698
DB 721 TGTCTCTCAGCCAGCCAGTATGATACCCCGAAGCTGGGAAGCAGGTCTACTGCTGCC 780
QY 699 ProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArg 718
DB 781 CCAGAGCAATTTGGTGTACCAAGTGCCTCCAGAGCAGCCCAAGAGTATCCGGAGCAGGAAG 840
QY 719 Asp---GlyGluCysAspValPheLysGlu 727
DB 841 GGATCGGGGAATGTGAATGGTTTAAAGAA 870

RESULT 2

BM905706

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..1061

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5556057"

/clone_lib="NIH_MGC_71"

/tissue_type="leiomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SpOrt6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

BASE COUNT 281 a 241 c 255 g 283 t

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

5,71e-123

1327.50

86.61%

83.33%

34.14%

14

Length:

Matches:

Conservative:

Mismatches:

Indels:

Gaps:

1061

280

11

25

34

7


```
Db 728 GCAGGGAAGTCAACACAGTGTGAGGCCGACTGTGACAGAAATATGACAGAGCCCTCGTG 787
QY 581 GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGlu 600
Db 788 GGAGGAATTTGGACACTCCATCAGATCCACTTAAACCTTGAGGAGGAGGAA 847
QY 601 AlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyr 620
Db 848 GCAGGCTATGCAGACCTCGATCTTACAACTCTCCAATGCAGGAAGTGTACACGCCAT 907
QY 621 AlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetAspMetSer 640
Db 908 GCTGAACCACTGCCGTAGGGGGCTGAGTACGCAACCCGATGCTATGACATGTCA 967
QY 641 GlyHisProThrSerValGlyGlnProSerThrPheLysAlaThrGlyAsn 660
Db 968 GGCACCCACAGCCTCAGTTGGTCTGCCCTCCACATCCACCTTCAAACTCGAGGACC 1027
QY 661 GlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSer 680
Db 1028 CAGCCTCACGCTTTAGTGGAACTTACACACTCTCTCCAGGACTGACAGCTGTTC 1087
QY 681 SerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAsp 700
Db 1088 TCAGGCCAGGCTCAGTATGACACCCCAAGGTGGGAG---TCAGCTGCTACCCACAG 1144
QY 701 GluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGly 720
Db 1145 GAAGTGGTATACAGTGGCCAGACGCCAGGAGCTATCAGGACGAGGAAGGATGATG 1204
QY 721 GluCysAspValPheLysGluIleLeu 729
Db 1205 AAGTTTGATGCTTTAAAGAAATCCTT 1231

RESULT 6
LOCUS A1435602/c
DEFINITION th79c10.x1 Soares_NHMPu.S1 Homo sapiens cDNA clone IMAGE:2124882
3' similar to TR:Q14089 Q14089 HYPOTHETICAL PROTEIN ;; mRNA
SEQUENCE.
ACCESSION A1435602.1 GI:4304334
VERSION EST.
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 641)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 865 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 450.
Location/Qualifiers
1..641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2124882"
/tissue_lib="Soares_NHMPu.S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="organ: mixed (see below); Vector: pTT3D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
```

```
BASE COUNT 161 a 125 c 153 g 202 t
ORIGIN
Alignment Scores: 2.1e-99 Length: 641
Pred. No.: Score: 1090.50 Matches: 209
Percent Similarity: 98.60% Conservative: 2
Best Local Similarity: 97.66% Mismatches: 3
Query Match: 28.05% Indels: 1
DB: Gaps: 0
US-10-060-830-3 (1-729) x A1435602 (1-641)
QY 295 AlaAlaPheAlaThrAspGluTyrGlnTyrLeuGlnIleAspLeuAsnLysGluLysLys 314
Db 641 TCTGCTTTTCCCTCTGATGAATACCAGTGTGTACAAATAGATTGTAAGGAAAGAAA 582
QY 315 IleThrGlyIleIleThrThrGlySerThrMetValGluHisAsnTyrTyrValSerAla 334
Db 581 ATAACAGGCATTATACCCACTGGAT--ACCATGGTGGAGCACAAATTACTATGTCTGCC 524
QY 335 TyrArgIleLeuTyrSerAspGlyGlnLysTyrThrValTyrArgGluProGlyVal 354
Db 523 TACAGAACTCTGTACAGTGTATGGCAGAAATGGACTGTGTACAGAGAGCCGTGTG 464
QY 355 GluGlnAspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspValArgAsnPhe 374
Db 463 GAGCAAGATAAGATATTCAGGAAACAAGATTATCACCAGGATGTCGTAATAACTTT 404
QY 375 LeuProProIleIleAlaArgPheIleArgValAsnProThrGlnTyrGlnLysIle 394
Db 403 TTGCCACCAATTTATTCACGCTTTTATACAGTGAATCCTACCAATGGCAGCAGAAAAT 344
QY 395 AlaMetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProLysLeu 414
Db 343 GCCATGAAATGGAGTGTCTGGATGTCTAGTTTATTCCTAAAGGTCTCTCCAAAACCT 284
QY 415 ThrGlnProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProLys 434
Db 293 ACTCAACTCTCCACTCTCCGAAACCAATGACCTCAAAACACTACAGCCCTCCAAA 224
QY 435 IleAlaLysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGlu 454
Db 223 ATAGCCAAAGGTCTGCCCCAAAATTTACGCAACCACTACAACCTCGCAGTAGCAATGAA 164
QY 455 PheProAlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThr 474
Db 163 TTTCCTGCACACAGACAGAAACAACTGCCAGTCTCTGATATCAGAAATACTTACCGTA 104
QY 475 ProAsnValThrLysAspValAlaLeuAlaValLeuValProValLeuValMetVal 494
Db 103 CCAATGTAAACCAAGATGATAGCGCTGGCTGAGTTCTTGTCCCTGTGCTGTCATGGTC 44
QY 495 LeuThrThrLeuIleLeuIleLeuValCysAlaTrpHisTrp 508
Db 43 CTCACACTCTCTCTCATATAGTGTGTGCTGTGGCACTGG 2

RESULT 7
LOCUS BQ014739/c
DEFINITION UI-H-ED1-act-n-12-0-UI-s1 NCI-CGAP_ED1 Homo sapiens cDNA clone
IMAGE:5833547 3', mRNA sequence.
ACCESSION BQ014739
VERSION BQ014739.1 GI:19739640
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 680)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES

Location/Qualifiers

1..680

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5833547"

/clone_lib="NCI-CGAP_ED1"

/tissue_type="Chondrosarcoma"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/note="Organ: Left Pubic Bone; Vector: p7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI-CGAP_ED1 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma cell line C95. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCTCAAGGCT.

TAG_LIB=UI-H-ED1

TAG_TISSUE=chondrosarcoma

TAG_SEQ=CGTCAGGCT"

BASE COUNT 138 a 157 c 172 g 213 t

ORIGIN

Alignment Scores:

Pred. No.: 5,15e-99 Length: 680

Score: 1087.00 Matches: 204

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 27.96% Indels: 0

Dbs: 14 Gaps: 0

US-10-060-830-3 (1-729) x BQ014739 (1-680)

QY 526 ArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGlu 545

Db 680 CGGCGAGTGTGGGAAGATGAAGCAGTTCCTCTGCAAAACGAGTGACCATGAG 621

QY 546 GluThrProValArgTyrSerSerGluValAsnHisLeuSerProArgGluValThr 565

Db 620 GAAACCCAGTTCGCTATAGCAGCAGCAGGAAGTAAATACCTGAGTCCAGAGAGAGTCCACC 561

QY 566 ThrValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGly 595

Db 560 ACAGTCTGCGAGCTGACTCTGAGAGTATGCTACGCCACTGGTAGGGAATTTGGT 501

QY 586 ThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGluAlaGlyTyrAlaAsp 605

Db 500 ACAGTTCATCAAGATCTACCTTTAAACAGAGAGAGAGAAAGCAGGCTATGAGAC 441

QY 606 LeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuPro 625

Db 440 CTAGATCCTTCAACTCACCAGGCGAGGAAGCTTTATCATGCTATGCTGAACCACTCCCA 381
QY 626 IleThrGlyProGluTyrAlaThrProIleleMetaspMetSerGlyHisProThrThr 645
Db 380 ATTACGGGGCGCTGAGTAGTCAACCCCAATCATGACATGTGAGGCGACCCCACTACT 321
QY 646 SerValGlyGlnProSerThrPheLysAlaThrGlyAsnGlnProProProLeu 665
Db 320 TCAGTTGTGTGAGCCCTCCATCCACTTTCAAGGCTACGGGAGCAACCTCCCACTA 261
QY 666 ValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAlaGln 685
Db 260 GTGGGAAGCTTCAATACACTTCTCTCCAGGAGTACAGCTGCTCTCAGCCAGCCAG 201
QY 686 TyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTyrGln 705
Db 200 TATGATACCCCAAGAGCTGGGAAGCCAGGTCTACCTGCCAGCAGCAATGGTGACCAG 141
QY 706 ValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGlyGlyCysAspValPhe 725
Db 140 GTGCCACAGACACACACAGAGATATCAGAGCAGGAGGATGGGGAATGTGATGTTT 81
QY 726 LysGluIleLeu 729
Db 80 AAAGAAATCCCTT 69
RESULT 8
BF732384
LOCUS
DEFINITION
693 bp mRNA linear EST 30-MAR-2001
naell12.x1 NCI-CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3434806 3', similar to TR:Q14089 Q14089 HYPOTHETICAL 40.0 KD PROTEIN ;, mRNA sequence.
ACCESSION BF732384
VERSION BF732384.1 GI:12057542
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 693)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 490.
Location/Qualifiers
1..693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3434806"
/clone_lib="NCI-CGAP_Ov18"
/tissue_type="fibrotheoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: p7T3-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTCGCGCGCGACACATTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization, and was

FEATURES

Location/Qualifiers

1..693

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3434806"

/clone_lib="NCI-CGAP_Ov18"

/tissue_type="fibrotheoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: p7T3-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTCGCGCGCGACACATTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization, and was

BASE COUNT	constructed by Bento Soares and M. Fatima Bonaldo.	
ORIGIN	162 a	197 t
	173 c	159 g
		2 others

Alignment Scores:	
Pred. No.:	1,728-97
Score:	1072.00
Percent Similarity:	99.02%
Best Local Similarity:	99.02%
Query Match:	27.5%
DB:	12
Length:	593
Matches:	203
Conservative:	0
Mismatches:	2
Indels:	0
Gaps:	0

US-10-060-830-3 (1-729) x BF732384 (1-693)

Qy	1	MetProLeuPheLeuLeuLeuValLeuLeuLeuLeuGluAspIaGlyAla	20
Db	79	ATGCCTGTTCCTCCTGCTTACTTGTCCTGCTCCTGCTCGAGACGCTGGAGCC	138
Qy	21	GlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSer	40
Db	139	CAGCAAGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGACACCCTTACATCC	198
Qy	41	IleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluLeuArgValLys	60
Db	199	ATAAACTACCCACAGACCTATCCCAACAGCACGTGTGTGAATGGGAGATCGCTGTAAG	258
Qy	61	MetGlyGluArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCys	80
Db	259	ATGGGAGAGAGATTCCCATCAAATTTGGTGACTTTGACATTGAAGATTCTGATCTCTGT	318
Qy	81	HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys	100
Db	319	CACITTAATTACTTGACAATTTATAATGGAAATGGAGTCAGCAGAACTGAAATAGSCAAA	378
Qy	101	TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerIysGlyAsnGluIleThr	120
Db	379	TACTGTGCTCGGGTTGCAATGAACCATTCAAATGAATCAAAAGGCAATGAATAATCA	438
Qy	121	LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal	140
Db	439	TTGCTGTTTCATGAGTGGAAATCCATGTTCTGAGCGCGGATTTTGGCCCTCATCTCTGT	498
Qy	141	IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu	160
Db	499	ATAGATAAACAAGATCTAAATTACTTGTGTGGACACTGCATCCCAATNTTTGGAAACCTGAG	558
Qy	161	PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIle	180
Db	559	TTCAAGTAGTACTGCCAGCTGGTGTCTGCTTNCITTTGCTCAGATATCTGGAAACAAT	618
Qy	181	ProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValVal	200
Db	619	CCTCATGGATATAGAGATTCTTCGCCATTGTGCATGGCTGGTGTGCATGCAGGAGTAGTG	678
Qy	201	SerAsnThrLeuGly	205
Db	679	TCAAACACAGTTGGGG	693

RESULT	9
BGI78371	
LOCUS	BGI78371
DEFINITION	60233014.f1 NIH MGC 91 Homo sapiens cDNA clone IMAGE:4431275 5'
	908 bp mRNA linear EST 06-FEB-2001

minna sequence.

ACCESSION	BG178371
VERSION	BG178371.1
KEYWORDS	EST.
SOURCE	human.

GI:12685074

SOURCE: Haman et al.
ORGANISM: *Homo sapiens*
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE: 1 (bases 1 to 908)
AUTHORS: NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.
TITLE: National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLAM10186 row: b column: 12
High quality sequence start: 2
High quality sequence stop: 654.

FEATURES	SOURCE
----------	--------

source

```
1. .908
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1="IMAGE:4431275"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="PH10B (phage-resistant)"
/notes="Organ: prostate; Vector: pCMV-SPORTS;
Site:2: Salt: Cloned unidirectionally; ol
Average insert size 1.4 kb. Library enric
full-length clones and constructed by Lif
Note: this is a NIH_MGC Library."
229 a 219 c 216 g 242 t 2 others
```

BASE COUNT	229 a	219 c	216 g	2 others
ORIGIN				

Alignment Scores:

Alignment scores:	
Pred. No.:	8,19e-97
Score:	1067.00
Percent Similarity:	77.99%
Best Local Similarity:	75.08%
Query Match:	72.44%
DB:	12
Gaps:	7
Indels:	38
Mismatches:	31
Conservative:	9
Matches:	232
Length:	908

US-10-060-830-3 (1-729) X BG178371 (1-908)

Qy	1	MetProLeuPheLeuLeuLeuLeuValLeuLeuLeuLeuLeuGluAspAlaGlyAla	20
Db	21	ATGGCTCTGTTCCTCCCTGCTTACTTGCTCGTCTGCTGCTCGAGACGCTGGAGCC	80
Qy	21	GlnGlnGlyAspGlyCysGlyHisThrValLeucGlyProGluSerGlyThrLeuThrSer	40
Db	81	CAGCAAGGTGATGATGGGCACACCTGTACTAGGCCCTGAGAGTGGAACCCCTTACATCC	140
Qy	41	IleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluIleArgValLys	60
Db	141	ATAAATACCACACAGACCTATCCCACACGACTGT-TGTGAATGGAGATCCGTGTAAAG	199
Qy	61	MetGlyGluArgValArgIleGlyPheGlyAspPheAspIleGluAspSerAspSerCys	80
Db	200	ATTGGCAGACAGAGTTCCGATCAAATTTTGGTGACTTTGCACATTGAAGATTCTGATTCCTGT	259
Qy	81	HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys	100
Db	260	CACITTAATTACTTGAGAATTTAATGAATTTGGAGTCACAGAACTGAAATAGGCCAAA	319
Qy	101	TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr	120
Db	320	TACTGTGCTCTGGGTGCCAAATCAACCATTCATTAATCAAAAGGCAATGAATACACA	379
Qy	121	LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal	140
Db	380	TGTGTTGTCATGAGTGGAAATCCATGTTTCTCGACGCGGATTTTTGGCCTCATACTCTGTT	439
Qy	141	IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu	160
Db	440	ATAGATAAACAGATCTAATTACTTTTGGACACTGCATCCATTTTGGAACTTGAG	499
Qy	161	PheSerLysTyrCysProAlaGlyCysLeuLeuPropHeAlaGluIleSerGlyThrIle	180

```

Db 500 TTCAGTAACTAGTCCAGCTGGTGTCTGCTCCTCTTCTGCTGAGATATCTGGACAAT 559
Qy 181 ProHisGlyTyrArgSerProLeuGlyValHisAlaGlyValVal 200
Db 560 CCTCATGATATAGAGATCTCTGCCATTTGTCATGCTGCTGTCATGACAGGAGTAGT 619
Qy 200 lSer---AsnThrLeuGlyGlnIleSerValValIleSerLysGlyIle---ProTyr 218
Db 620 GTCACACACACNGTTGGGGCGGCAAGTCAAGTGTGTAAATCAGCTACAGGTATCCCTTTA 679
Qy 218 r-TyrGluSerSerLeu-AlaAsnVal---ThrSerValValGly-----HisLeu 234
Db 680 TTAATGAAGTTCCTTGGGTCAACAGCTTCACACTCTGTGGGGGGGACACTTTTCTA 739
Qy 235 SerThrSerLeuPheThrPheLysThrSerGlyCystyrGlyThrLeuGlyMetGluSer 254
Db 740 CAACGTCTCTTCACACTTTTCACACACACGGCGCTTC----- 778
Qy 255 GlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluTyrThrAspHisThr 274
Db 779 ----- 793
Qy 275 Gly---GlnGluAsnSerTrpLysProLysLysAlaArg-----Leu 287
Db 794 GGACCCGGGGGATAGCCGGGACCCCAAAAAACACACACCATGGGGGGGATGTCAGTC 853
Qy 288 LysLysProGlyProTrp 294
Db 854 CCCACCGGGCACACCTGG 874

RESULT 10
BE894226 601437167F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922169 5',
LOCUS mRNA sequence.
DEFINITION BE894226 618 bp mRNA linear EST 20-OCT-2000
VERSION BE894226
KEYWORDS EST.
SOURCE BE894226.1 GI:10356380
ORGANISM human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 618)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9756 row: e column: 18
High quality sequence stop: 618.
FEATURES
Location/Qualifiers
1..618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3922169"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notice="Organ: skin; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 201 a 135 c 139 g 143 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 9.3le-97 Length: 618
Score: 1064.00 Matches: 202
Percent Similarity: 98.06% Conservative: 0
Best Local Similarity: 98.06% Mismatches: 3
Query Match: 27.37% Indels: 2
DB: 12 Gaps: 0

US-10-060-830-3 (1-729) x BE894226 (1-618)
Qy 237 SerLeuPheThrPheLysThrSerGlyCystyrGlyThrLeuGlyMetGluSerGlyVal 256
Db 3 AGTCTTTTACATTTAGACACAGTGGATGTTATGACACATGGGGATGGAGTCTGGTGTG 62
Qy 257 lIleAlaAspProGlnIleThrAlaSerSerValLeuGluTyrThrAspHisThrGlyGln 276
Db 63 ATCGGGATCTCTCAATAACAGCATCATCTGTCTGGAGTGGAGTACACACAGGGCAA 122
Qy 277 GluAsnSerTrpLysProLysLysAlaArgLeuLysLysProGlyProTrpAlaAla 296
Db 123 GAGAACAGTTTGGAAACCCAAAAAGCCAGGCTGAAAAACCTGGACCCCTTGGGCTGCT 182
Qy 297 PheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThr 316
Db 183 TTTGCCACTGATGATACCATCCAGTGGTTACAAATAGATTGGAATAGGAAAAAGAAATACAA 242
Qy 317 GlyIleIleThrThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArg 336
Db 243 GGCATTATACCACTGGATCCACCATGGTGAGGACACATTTACTATGTCTGCTGCTACAGA 302
Qy 337 lIleLeuTyrSerAspAspGlyGlnLysTrpThrValTyrArgGluProGlyValGluGln 356
Db 303 ATCCTGTACAGTGTATGATGGGAGAAATGGAGTGTGTACAGAGAGCCCTGGTGTGGAGCAA 362
Qy 357 AspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspValArgAsnAspPheLeuPro 376
Db 363 GATAAGATATTTCAGAGAAACAAAGATTATCACCAGGATGTGCGTAATAAATTTTGGCA 422
Qy 377 -ProIleIleAlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMe 396
Db 423 CCCAATATTACACGTTTATTATAGATGAATCCTACCCATGGCAGCAGAAATTTGCCAT 482
Qy 396 lLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGln 416
Db 483 GAAATGGAGCTGCTCGGATGTCAGTTTATTCCTAAAGTGTCTCTCCAAAATTTACTCA 542
Qy 416 nProProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAl 436
Db 543 ACT-CCACCTCTCGGAACACAGCAATGACCTCAAAACACTACAGCCCTCCAAAATAGC 601
Qy 436 aLysGlyArgAlaPro 441
Db 602 CAAGGGTCTGTGCC 617

RESULT 11
AA147037/zc 671 bp mRNA linear EST 05-DEC-1996
LOCUS z032a05.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
DEFINITION IMAGE:588560 3' similar to TR:G704441 G704441 HYPOTHETICAL 40.0 KD
PROTEIN,, mRNA sequence.
AA147037 GI:1716444
VERSION AA147037.1 EST.
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 671)
REFERENCE Hillier, L., Lennon, G., Becker, M., Bonaudo, M.F., Chisapelli, B.,
AUTHORS Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins
, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Le, N., Mardis, E., Moore
, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.

```

TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 9704478
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (infoimage.llnl.gov) for further information.
 Seq primer: -40M13 fwd. from Amersham
 High quality sequence stop: 439.

FEATURES

Location/Qualifiers
 1..671
 /organism="Homo sapiens"
 /db_xref="GDB:460862"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:588560"
 /clone_lib="Stratagene colon (#937204)"
 /tissue_type="tumor"
 /cell_line="T84 carcinoma cell line"
 /lab_host="SOUR cells (kanamycin resistant)"
 /note="Organ: colon; Vector: pBluescript SK-; Site: 1:
 EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. T-84 colonic epithelial cell line. Average
 insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
 sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5'
 CTCGACGTTTTTTTTTTTTTTT 3"

BASE COUNT 176 a 162 c 133 g 198 t 2 others
 ORIGIN

Alignment Scores:
 Pred. No.: 4,84e-95 Length: 671
 Score: 1047.50 Matches: 210
 Percent Similarity: 95.89% Conservative: 0
 Best Local Similarity: 95.89% Mismatches: 8
 Query Match: 26.94% Indels: 4
 DB: 9 Gaps: 1

US-10-060-830-3 (1-729) x AAL47037 (1-671)

QY 173 PheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerProLeuCysMet 192
 Db 651 TTTGCTGAGATATCTGGACC-ATTCTCATGTGATATGA-GATTCTCTGCCCA--TGTCNT 597
 QY 193 AlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValValIle 212
 Db 596 GGCTGGTGTATGTCAGGAGTAGTGCAACACAGTT-GCGGCCCAATCAGTGTGTAATT 538
 QY 213 SerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThrSerValValGly 232
 Db 537 ACTAAGGTATTCCTATATGAAAGTCTTTTAAAGACAAGTGCATGTTATGGAACACTGGGATG 478
 QY 233 HisLeuSerThrSerLeuPheThrPhelYsThrSerGlyCysTyrGlyThrLeuGlyMet 252
 Db 477 CACTTATCTACAAGTCTTTTACATTAAAGACAAGTGCATGTTATGGAACACTGGGATG 418
 QY 253 GluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluTrpThrAsp 272
 Db 417 GAGTCTGGTGTATCGCGGATCTCTCAATAACAGCATCATCTGCTGGAGTGGACTGAC 358
 QY 273 HisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLysLysProGlyPro 292
 Db 357 CACACAGGCAAGAACAGATTGGAAACCCAAAAAGCCAGCGCTGAAAAACCTGGACCC 298
 QY 293 ProTyrAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGlu 312
 Db 297 CCTGGGCTGCTTTTGGCCACTGATGAATACCAAGTGGTTACAAATAGATTGTAAGAAGAA 238
 QY 313 LysLysIleThrGlyIleIleThrThrGlySerThrMetValGluHisAsnTyrTyrVal 332
 Db 237 AAGAAATAACAGGCATTATACCACTGGATCACCATTGGAATCAGTGGTGGAGCACAACTACTATG 178

QY 333 SerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTyrThrValTyrArgGluPro 352
 Db 177 TCTGCCACAGATCCCTGTCAGTGTATGGCAGAAATGGACTGTGTACAGAGAGCT 118
 QY 353 GlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspValArgAsn 372
 Db 117 GGTGTGAGCAAGATAAGATATTTCAGGAACAAGATTATCACCAGGATGTCGTAAT 58
 QY 373 AsnPheLeuProPheIleAlaArgPheIleArgValAsnProThrGlnTrpGln 391
 Db 57 AACTTTTGGCCACCAATTATTCACGTTTATTAGTGAATCTCTACCAATGTCGAC 1

RESULT 12
 BB179946
 LOCUS BB179946 643 bp mRNA linear EST 19-OCT-2001
 DEFINITION musculus cDNA clone A230081I02 3', mRNA sequence.
 ACCESSION BB179946
 VERSION BB179946.2 GI:16269978
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 643)
 AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
 M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
 D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 Unpublished (2001)
 JOURNAL On Jun 29, 2000 this sequence version replaced gi:8839029.
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,
 K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
 Hayashizaki,Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 00:35:34 ; Search time 5021.22 Seconds
(without alignments)
11371.681 Million cell updates/sec

Title: US-10-060-830-1113
Perfect score: 1962
Sequence: 1 atgcctctgtctctctgtct.....tttttaagaatacctttga 1962

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1897.4	96.7	2939	9	AF387547 Homo sapi
2	1895.8	96.6	6093	9	AB073146 Homo sapi
3	1842.2	93.9	3858	9	BC029658 Homo sapi
4	1440.2	73.4	2339	10	AF387549 Rattus no
5	1433.8	73.1	2642	10	AF387548 Mus muscu
6	1230.2	62.7	2020	9	AK056350 Homo sapi
7	1159.2	59.1	1907	6	AX405795 Sequence
8	932.8	47.5	1388	9	HUMCUB1
9	473.8	24.1	192247	2	AC013497 Homo sapi
10	473.8	24.1	192247	2	AC013497 Homo sapi
11	462.8	23.6	146515	2	AC025661 Homo sapi
12	353.8	18.0	203102	2	AC126975 Rattus no
13	218	11.1	192247	2	AC013497 Homo sapi
14	204	10.4	1464	6	AX118816 Sequence
15	204	10.4	1620	6	AX118820 Sequence
16	204	10.4	1761	6	AX118818 Sequence
17	204	10.4	1768	6	AX118822 Sequence
18	204	10.4	2547	9	AK095973 Homo sapi
19	174.8	8.9	217	6	AX182030 Sequence
20	141.8	7.2	586	9	BC007117 Homo sapi
21	141.8	7.2	164216	2	AC091213 Homo sapi
22	135.4	6.9	146515	2	AC025661 Homo sapi
23	135.2	6.9	59871	2	AC101046 Mus muscu
24	131.2	6.7	61344	2	AC073443 Homo sapi
25	130.4	6.6	203102	2	AC126975 Rattus no
26	119	6.1	2845	10	BC026771 Mus muscu
27	94.4	4.8	6895	4	BOVFA2V2A
28	94.4	4.8	6910	4	BOVFA2V2A
29	92.8	4.7	1895	4	S80643 Bos taurus
30	92.8	4.7	1965	4	BTPAS67PT
31	92.8	4.7	59871	2	AC101046 Mus muscu
32	91.6	4.7	6585	10	MMU52925
33	91.6	4.7	7062	4	AF191308 Sus scrof
34	90.6	4.6	61344	2	AC073443 Homo sapi
35	89.8	4.6	1719	9	HSU70312 Homo sapien
36	89.8	4.6	2728	9	BC030828 Homo sapi
37	89.8	4.6	4681	6	AX473349 Sequence
38	88.8	4.5	1227	9	BTBP47PRO
39	88.2	4.5	6893	9	HUMFVA
40	88	4.5	2303	10	AF031524 Mus muscu
41	86.6	4.4	6909	6	A46255 Sequence 1
42	86.6	4.4	6909	6	A63218 Sequence 1
43	86.6	4.4	6909	6	AF071379 Sequence
44	86.6	4.4	6909	6	AX146885 Sequence
45	86.6	4.4	6909	6	AX409628 Sequence

ALIGNMENTS

RESULT 1
AF387547
LOCUS AF387547
DEFINITION Homo sapiens endothelial and smooth muscle cell-derived
neuropilin-like protein (ESDN) mRNA, complete cds.
ACCESSION AF387547
VERSION AF387547.1 GI:16902434
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2939)
AUTHORS Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N.,
Matsumori,A., Sasayama,S., Honjo,T. and Tashiro,K.

TITLE ESDN, a novel neuropilin-like membrane protein cloned from vascular cells with the longest secretory signal sequence among eukaryotes, is up-regulated after vascular injury
J. Biol. Chem. 276 (36), 34105-34114 (2001)
J. Biol. Chem. 276 (36), 34105-34114 (2001)
PUBMED 21424027
REFERENCE 11447234
AUTHORS Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N., Sasayama, S., Matsumori, A., Honjo, T. and Tashiro, K.
TITLE Direct Submission
JOURNAL Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan
FEATURES Location/Qualifiers
source
1. .2939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
gene
1. .2939
/gene="ESDN"
CDS
12. .2339
/gene="ESDN"
/codon_start=1
/product="endothelial and smooth muscle cell-derived neuropilin-like protein"
/protein_id="AAL30178.1"
/db_xref="GI:16902435"
translation="MASRAVARRRCQPCQVRAAAAPAAWALFLSRSLPCCSSSS
ESMLPALLLVLLLLDAGACGGHVLGPESTLTLSINPOTYPNSTVCEWE
IRVMGERVIRKFGDFIEDSDSCHFYRLIYNGIVSRTEIGKYGLGLOMNHIES
KNEITLFFSGHVSGLASVYIDKDLITLCLDTASNFLEPERSKYCPACCLLP
FAEISGTHPHRSPLCMAGVHGVNSLTGGISVVISKGIPIYESSLIANNVTSV
VHLSLTSLFTKSGCYGLNMGESVADPOITASVLEWTDHTGOENSMKPKARLK
KGPMPAAFAFDEYOWLOIDLNKKITGITTTGSTVMHNYVYSAVRLYSDGQKW
TYRPFGEQDKIFGKNKDYHQDVRRNLFPLIAEFIRVNPTQOQKAMKWEILLGCQ
FIPGRPRPLTTPPPRRNSDLKNTTAPPKIAGRPKRFPTOPLOPRSNNEPQAEQT
TASPDRIKNTVPTKDVALAALVPLVLMVLTLLILLYCAHNRNRKKTETGYD
LPYWRAGWMKGMKFLPAKVADEETEVYSSSEVNHLSPREVTYVLDQSAEYAP
LVGGVLGTLHORSTFKPEEGEAGVADLPYNSPQOEYHAYAEPLPTTGPEYATPII
XDMSPHTPSVGPSTSTFKATGNOPPLVGTNTLLSRDSCSSAQAOYDTPRAGKP
GUPADELYVQVQSTQSVSGNRRGECDFKEIL"
BASE COUNT 825 a 675 c 657 g 782 t
ORIGIN
Query Match 96.7%; Score 1897.4; DB 9; Length 2939;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1901; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 56 GAGCCGACGAAGCAATCTGTGGTGGGTGGCAATGAACCAATCAATGAATCAA 115
DB 433 GAACGTAATAGCAATATCTGTGGTGGGTGGCAATGAACCAATCAATGAATCAA 492
QY 116 AAGCAATGAATACATCTGTTGATGAGTGGATCCATGTTCTGGACGGGATTTT 175
DB 493 AAGCAATGAATACATCTGTTGATGAGTGGATCCATGTTCTGGACGGGATTTT 552
QY 176 TGGCTCATCTCTGTTATAGATAAACAAGATCTAATCTGTTGGACACTGCATCCA 235
DB 553 TGGCTCATCTCTGTTATAGATAAACAAGATCTAATCTGTTGGACACTGCATCCA 612
QY 236 ATTTTGGACCTGAGTTCAGTAAGTACCTGCCAGCTGGTGTCTCTCTCTTTTCTG 295
DB 613 ATTTTGGACCTGAGTTCAGTAAGTACCTGCCAGCTGGTGTCTCTCTCTTTTCTG 672
QY 296 AGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTGCCATTTGTGATGGCTGGT 355
DB 673 AGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTGCCATTTGTGATGGCTGGT 732
QY 356 TGCATGAGGAGTAGTGTCAACACAGTTGGCGGCCAAATCAGTGTGTAATAGTAAG 415
DB 733 TGCATGAGGAGTAGTGTCAACACAGTTGGCGGCCAAATCAGTGTGTAATAGTAAG 792
QY 416 GTATTCCTTATATGAAAGTCTTTGGCTAACACAGTCACATCTGTGTGGGACACTTAT 475

DB 793 GTATTCCTTATATGAAAGTCTTTGGCTAACACGTCACATCTGTGTGGGACACTTAT 852
QY 476 CTACAAGTCTTTTACATTTTAAGACAAGTGTATTTAGAACACTGGGATGAGTCTG 535
DB 853 CTACAAGTCTTTTACATTTTAAGACAAGTGTATTTAGAACACTGGGATGAGTCTG 912
QY 536 GTGTGATCGGGATCCCTCAATAACACGATCATCTGTCTGGAGTGGACTGACCACACAG 595
DB 913 GTGTGATCGGGATCCCTCAATAACACGATCATCTGTCTGGAGTGGACTGACCACACAG 972
QY 596 GGCAAGAGACAGTTGGAAACCCCAAAAGCCAGGCTGAAAAAACCTGGACGCCCTTGGG 655
DB 973 GGCAAGAGACAGTTGGAAACCCCAAAAGCCAGGCTGAAAAAACCTGGACGCCCTTGGG 1032
QY 656 CTGCTTTGGCCACTGATGAATACCAAGTGGTTACAAATAGATTGTAAGAGAAAGAAA 715
DB 1033 CTGCTTTGGCCACTGATGAATACCAAGTGGTTACAAATAGATTGTAAGAGAAAGAAA 1092
QY 716 TAACAGGCATTATTAACCACTGGATCCACATGGTGGAGCACAAATTACTATGTGTGCT 775
DB 1093 TAACAGGCATTATTAACCACTGGATCCACATGGTGGAGCACAAATTACTATGTGTGCT 1152
QY 776 ACAGAACTCTGTACAGTGTGATGGGAGAAATGAGACTGTGTACAGAGAGCCCTGTGTGG 835
DB 1153 ACAGAACTCTGTACAGTGTGATGGGAGAAATGAGACTGTGTACAGAGAGCCCTGTGTGG 1212
QY 836 AGCAAGATAAGATATTTCAAGAAACAAAGATTATCACCAGAGTGTGCGTAATAACTTTT 895
DB 1213 AGCAAGATAAGATATTTCAAGAAACAAAGATTATCACCAGAGTGTGCGTAATAACTTTT 1272
QY 896 TGCCACCAATATTTGACAGTGTGATGGGAGAAATGAGACTGTGTACAGAGAGCCCTGTGTGG 955
DB 1273 TGCCACCAATATTTGACAGTGTGATGGGAGAAATGAGACTGTGTACAGAGAGCCCTGTGTGG 1332
QY 956 CCATGAAATAGGAGTGTGCTGGATGTGCTGATTTATTTCTTAAAGTGTGCTGTCCCAAACTTA 1015
DB 1333 CCATGAAATAGGAGTGTGCTGGATGTGCTGATTTATTTCTTAAAGTGTGCTGTCCCAAACTTA 1392
QY 1016 CTCAACCTCCACTCTCGGACAGCAATGACCTCAAAACACTACAGCCCTCCCAAAA 1075
DB 1393 CTCAACCTCCACTCTCGGACAGCAATGACCTCAAAACACTACAGCCCTCCCAAAA 1452
QY 1076 TAGCCAAAGTGTGCTGCCCAAAATTTACGCAACCACTTACCACTCGCAGTAGCAATGAAT 1135
DB 1453 TAGCCAAAGTGTGCTGCCCAAAATTTACGCAACCACTTACCACTCGCAGTAGCAATGAAT 1512
QY 1136 TTCTCTCACAGACAGAACAACTGCCAGTCTCTGATATCAGAAATACACCGTAACCT 1195
DB 1513 TTCTCTCACAGACAGAACAACTGCCAGTCTCTGATATCAGAAATACACCGTAACCT 1572
QY 1196 CAAATTAACCAAGATGTAGCGCTGGCTGCAGTCTTCTTGTCCCTGTGCTGTGCTGCTCC 1255
DB 1573 CAAATTAACCAAGATGTAGCGCTGGCTGCAGTCTTCTTGTCCCTGTGCTGTGCTGCTCC 1632
QY 1256 TCACCTACTCTCATCTCATATTAGTGTGCTTGGCACTGGAGAAACAGAAAGAAAAA 1315
DB 1633 TCACCTACTCTCATCTCATATTAGTGTGCTTGGCACTGGAGAAACAGAAAGAAAAA 1692
QY 1316 CTGAAGGCCTATGACTTACCTTACTGGACCGGAGGTGGTGGAAAGGAATGAAGC 1375
DB 1693 CTGAAGGCCTATGACTTACCTTACTGGACCGGAGGTGGTGGAAAGGAATGAAGC 1752
QY 1376 AGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAAACCCAGTTTGGTGTATAGCAGCAGCG 1435
DB 1753 AGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAAACCCAGTTTGGTGTATAGCAGCAGCG 1812
QY 1436 AAGTTAATACCTGAGTCCAGAGAAAGTCCACAGTGTGCTGACGCTGACCTGTCAGAGT 1495
DB 1813 AAGTTAATACCTGAGTCCAGAGAAAGTCCACAGTGTGCTGACGCTGACCTGTCAGAGT 1872
QY 1496 ATGCTCAGCCACTGGTAGGAGAAATTTGTTGTACACTTCAATCAAGATCTACCTTTAAAC 1555

```
Db 1873 ATGCTAGCCACTGGTAGGGAATGTTGGTACACTTCATCAAGATCTACCTTTAAAC 1932
Qy 1356 CAGAAGAGAAAGAACAGGCTATGCAGACCTAGATCTCTTCAACTCACCAGGCGAG 1615
Db 1933 CAGAGAGAGAAAGAACAGGCTATGCAGACCTAGATCTCTTCAACTCACCAGGCGAG 1992
Qy 1616 AAGTTTATCATGCTGCTGAACCACTCCCAATACGGGGCCCTGAGATGATCAACCCCAA 1675
Db 1993 AAGTTTATCATGCTGCTGAACCACTCCCAATACGGGGCCCTGAGTATGATCAACCCCAA 2052
Qy 1676 TCATCATGGACATGCTAGGCGACCCCACTTCACTTGGTCAGCCCTCCACATCCACTT 1735
Db 2053 TCATCATGGACATGCTAGGCGACCCCACTTCACTTGGTCAGCCCTCCACATCCACTT 2112
Qy 1736 TCAAGGCTACGGGAACCACTCCGCCACTAGTGGGAACCTTCAATACACTTCTCTCCA 1795
Db 2113 TCAAGGCTACGGGAACCACTCCGCCACTAGTGGGAACCTTCAATACACTTCTCTCCA 2172
Qy 1796 GGACTGACAGTGTCTCTCAGCCAGGCCAGTATGATPACCCCGAAAGCTGGGAAGCCAG 1855
Db 2173 GGACTGACAGTGTCTCTCAGCCAGGCCAGTATGATPACCCCGAAAGCTGGGAAGCCAG 2232
Qy 1856 GTCTACCTGCCAGAGCAATGTTGTACAGGTGCCAGACAGACACAGAAGTATCAG 1915
Db 2233 GTCTACCTGCCAGAGCAATGTTGTACAGGTGCCAGACAGACACAGAAGTATCAG 2292
Qy 1916 GAGCAGGAAGGATGGGAATGTGATGTTTTAAAGAAATCCTTTGA 1962
Db 2293 GAGCAGGAAGGATGGGAATGTGATGTTTTAAAGAAATCCTTTGA 2339

RESULT 2
AB073146 Homo sapiens mRNA for CLCP1, complete cds. linear PRI 26-APR-2002
LOCUS AB073146
DEFINITION Homo sapiens mRNA for CLCP1, complete cds.
ACCESSION AB073146
VERSION AB073146.1 GI:20330503
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Koshikawa,K., Osada,H., Kozaki,K., Konishi,H., Masuda,A.,
Tatematsu,Y., Mitsudomi,T., Nakao,A. and Takahashi,T.
TITLE Significant up-regulation of a novel gene, CLCP1, in a highly
metastatic lung cancer subline as well as in lung cancers in vivo
JOURNAL Oncogene (2002) In press
REFERENCE
AUTHORS Koshikawa,K., Konishi,H., Osada,H. and Takahashi,T.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2001) Takashi Takahashi, Aichi Cancer Center
Research Institute, Division of Molecular Oncology, Chikusa-ku,
Kanokoden 1-1, Nagoya, Aichi 464-8681, Japan
(E-mail:takeaichi-cs.jp, Tel:81-52-764-2993, Fax:81-52-764-2993)
LOCATION/Qualifiers
FEATURES
source
1..6093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3q12.1-12.2"
/cell_line="normal lung tissue"
/tissue_type="lung"
1..6093
/gene="CLCP1"
/364..2691
/gene="CLCP1"
/function="associated with metastatic phenotype"
/note="CUB, LCCL, coagulation factor V/VIII-homology
domains protein 1"
/codon_start=1
/evidence="experimental"
/product="CLCP1"

/misc_feature
577..924
/gene="CLCP1"
/note="CUB domain"
934..1191
/gene="CLCP1"
/note="LCCL domain"
1234..1710
/gene="CLCP1"
/note="F5/8 type C domain"
1942..2010
/gene="CLCP1"
/note="transmembrane domain"
4264..4269
/gene="CLCP1"
5873..5878
/gene="CLCP1"
6052..6057
/gene="CLCP1"
BASE COUNT 1671 a 1258 c 1289 g 1874 t 1 others
ORIGIN
Query Match 96.6%; Score 1895.8; DB 9; Length 6093;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1900; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 56 GAGCCCAAGGCAAAATCTGTGTCTGGGTGCAAAATGAACCAATCAATTGAATCAA 115
Db 785 GAACCTGAATAGGCAAAATCTGTGTCTGGGTGCAAAATGAACCAATCAATTGAATCAA 844
Qy 116 AAGGCAATGAATACATCTGTGTCTGAGTGGAAATCCATCTTCTGACCGGATTTT 175
Db 845 AAGGCAATGAATACATCTGTGTCTGAGTGGAAATCCATCTTCTGACCGGATTTT 904
Qy 176 TGGCCTCATCTCTCTTATAGATAAACAAGATCTAATTACTTGTGTTGGACATGCATCCA 235
Db 905 TGGCCTCATCTCTCTTATAGATAAACAAGATCTAATTACTTGTGTTGGACATGCATCCA 964
Qy 236 ATTTTGGAACTGAGTTCAGTAACTAGTACTGCCAGCTGGTGTCTGCTCTCTTTGGTG 295
Db 965 ATTTTGGAACTGAGTTCAGTAACTAGTACTGCCAGCTGGTGTCTGCTCTCTTTGGTG 1024
Qy 296 AGATATCTGGAACAATTCCTCATGATATAGAGATTCCTCGCCATTTGTCATGGCTGTG 355
Db 1025 AGATATCTGGAACAATTCCTCATGATATAGAGATTCCTCGCCATTTGTCATGGCTGTG 1084
Qy 356 TGCATGCAGGAGTAGTGTCAAAACAGCTTGGGGCGGCAAAATCAGTGTGTTAATTAGTAAG 415
Db 1085 TGCATGCAGGAGTAGTGTCAAAACAGCTTGGGGCGGCAAAATCAGTGTGTTAATTAGTAAG 1144
Qy 416 GTATTCCTTATTAAGAACTTCTTTGGCTTAACAAGCTACATCTGTTGGTGACACTTAT 475
Db 1145 GTATTCCTTATTAAGAACTTCTTTGGCTTAACAAGCTACATCTGTTGGTGACACTTAT 1204
Qy 476 CTACAGTCTTTTACATTTAAGACAAGTGTATGTAACACTGGGGAGTGGAGTCTG 535
Db 1205 CTACAGTCTTTTACATTTAAGACAAGTGTATGTAACACTGGGGAGTGGAGTCTG 1264
```

```
/protein_id="BAB91138.1"
/db_xref="GI:20330504"
/translat="MASRAVRRARPCQPVRAAXAAPAAALPLSRSLPPCSNSS
FSPLFLILLVLLLEDAQAQDGCCHTILGSPESGLTISNYPTYPNVCWE
IRYKMERVIRKFGDFIEDSDCHNRYLRIYNGIGVSRTEIGYGLGLQMHSTES
KNEITLLENSGIHVSGRGFLASVVIDKODLITCLDTASNELEPEFSKYCPAGCLLP
FAGISGTHPGYRDSPLCMAGHAGVSVNTLGGQISVLSKGIPIYESSLANNTSV
VGHLSLFTFTSGCYTGLMESGVADPQITASSVLEWTDHTGQENSWKPKARLK
KPGPPAAAFATDEYQWLIDLNKEKITGITTTGSTMVEHNYTVSARYILYSDDGKQ
TVYREPGEQDKIFQGNKYHQVRRNLFPIIARFIRVNPQTQOQRIAMKMLLGCQ
FIPKRPPLKTOPPPRNSDLNKTAPPKIAKGRAPKFTQLOPQRSNFFPAQTET
TASPDIRNTTVPNTKDVALAALVPLVYMLVLTLLILVCAWHNRNKKTEGTV
LPWDRAGWKKMKQFLPAKAYDHEETPVYSSSEVNHLSPREVTTVLQADSAEYAP
LVGIGVTGTHQSTFTFKPEKGEAGIADLPYNSPGQEVYHAYAEPLTGPETATPII
MDMSGHPHTSVGPQSTSTFKATQNPPLVGYNTLLSRTDSCSSAAQAYDTPKAGKP
GLPAPDELVYVQSTQEVSGAGRDGECDFVEKIL"
577..924
/misc_feature
/gene="CLCP1"
/note="CUB domain"
934..1191
/misc_feature
/gene="CLCP1"
/note="LCCL domain"
1234..1710
/misc_feature
/gene="CLCP1"
/note="F5/8 type C domain"
1942..2010
/misc_feature
/gene="CLCP1"
/note="transmembrane domain"
4264..4269
/polyA_signal
/gene="CLCP1"
5873..5878
/polyA_signal
/gene="CLCP1"
6052..6057
/polyA_signal
/gene="CLCP1"
BASE COUNT 1671 a 1258 c 1289 g 1874 t 1 others
ORIGIN
```

Qy 536 GTGTGATCGCGATCTCTCAATAACAGCATCATCTGTCTGGAGTGAGCTGACACACAG 595
Db 1265 GTGTGATCGCGATCTCTCAATAACAGCATCATCTGTCTGGAGTGAGCTGACACACAG 1324
Qy 596 GGCAGAGAACAGTGTGAAACCCCAAAAGCCAGGCTGAAAAACCTGGACCGCTTGGG 655
Db 1325 GGCAGAGAACAGTGTGAAACCCCAAAAGCCAGGCTGAAAAACCTGGACCGCTTGGG 1384
Qy 656 CTGCTTTTCCACATGATCAATACAGTGTGTACAAATAGATTGAATAAGAAAGAAA 715
Db 1385 CTGCTTTTCCACATGATCAATACAGTGTGTACAAATAGATTGAATAAGAAAGAAA 1444
Qy 716 TAACAGGCATTATAACCACTGGATCCACCATGGTGGAGCACAATTAATGTGTCTGCCT 775
Db 1445 TAACAGGCATTATAACCACTGGATCCACCATGGTGGAGCACAATTAATGTGTCTGCCT 1504
Qy 776 ACAGAACTCTACAGTATGATGGGAGAAATGGATGTGTACAGAGAGCGCTGGTGTGG 835
Db 1505 ACAGAACTCTACAGTATGATGGGAGAAATGGATGTGTACAGAGAGCGCTGGTGTGG 1564
Qy 836 AGCAAGATAAGATATTTCAAGAAACAAAGATTATCACAGGATGCGTAAATACTTTT 895
Db 1565 AGCAAGATAAGATATTTCAAGAAACAAAGATTATCACAGGATGCGTAAATACTTTT 1624
Qy 896 TGCCACCAATTATTGCAGCTTTTATTAGAGTGAATCTTACCAATGGCAGCAGAAATTTG 955
Db 1625 TGCCACCAATTATTGCAGCTTTTATTAGAGTGAATCTTACCAATGGCAGCAGAAATTTG 1684
Qy 956 CCATGAAATGGAGCTGCTGGATGTGAGTTTATCTTAAAGTGTGCTTCCAAACTTA 1015
Db 1685 CCATGAAATGGAGCTGCTGGATGTGAGTTTATCTTAAAGTGTGCTTCCAAACTTA 1744
Qy 1016 CTCACCTCCACCTCTCGGAGCAGCAATGACCTCAAAACACTACAGCCCTCCAAAA 1075
Db 1745 CTCACCTCCACCTCTCGGAGCAGCAATGACCTCAAAACACTACAGCCCTCCAAAA 1804
Qy 1076 TAGCAAGAGTGTGCTGCGCCCAAAATTTAGCAACCACTACACCTCGCAGTAGCAATGAAT 1135
Db 1805 TAGCAAGAGTGTGCTGCGCCCAAAATTTAGCAACCACTACACCTCGCAGTAGCAATGAAT 1864
Qy 1136 TTCCTGCACAGACAGAACAAACACTGCGAGTCTGTATATCAGAAATACTACCGTAACCT 1195
Db 1865 TTCCTGCACAGACAGAACAAACACTGCGAGTCTGTATATCAGAAATACTACCGTAACCT 1924
Qy 1196 CAATGTATCAACCAAGATGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1255
Db 1925 CAATGTATCAACCAAGATGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1984
Qy 1256 TCACCTACTCTCATCTCTATATTAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1315
Db 1985 TCACCTACTCTCATCTCTATATTAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2044
Qy 1316 CTGAAGGCACCTATGACTTACCTTACTTGGGACCGGCGAGTGTGTTGAAAGGAATGAAGC 1375
Db 2045 CTGAAGGCACCTATGACTTACCTTACTTGGGACCGGCGAGTGTGTTGAAAGGAATGAAGC 2104
Qy 1376 AGTTTCTCTCTGCAAAAGCAGTGAACATGAGGAAACCCAGTTCGGTATAGCAGCAGCG 1435
Db 2105 AGTTTCTCTCTGCAAAAGCAGTGAACATGAGGAAACCCAGTTCGGTATAGCAGCAGCG 2164
Qy 1436 AAGTTAATCACTGATGCCAAGAGAGTCAACACAGTGTGCTGAGCTGACCTGCTGAGAGT 1495
Db 2165 AAGTTAATCACTGATGCCAAGAGAGTCAACACAGTGTGCTGAGCTGACCTGCTGAGAGT 2224
Qy 1496 ATGCTCAGCCACTGGTAGGAGGAATTTGTTGTACACTTCTATCAAGATCTACCTTTAAAC 1555
Db 2225 ATGCTCAGCCACTGGTAGGAGGAATTTGTTGTACACTTCTATCAAGATCTACCTTTAAAC 2284
Qy 1556 CAGAAGAGGAAAGAGCAGGCTATGACAGACCTTAGTTCCTTACAACTACACAGGCGAGG 1615
Db 2285 CAGAAGAGGAAAGAGCAGGCTATGACAGACCTTAGTTCCTTACAACTACACAGGCGAGG 2344
Qy 1616 AAGTTTATCATGCTATGCTGAACCACTCCCAATTTACGGGGCTGTAGTATGCAACCCCAA 1675

Db 2345 AAGTTTATCATGCTATGCTGAACCACTCCCAATTTACGGGGCTGTAGTATGCAACCCCAA 2404
Qy 1676 TCATCATGGACATGTTCAGGGACCCCAACACTTCAGTTGGTTCACCTCCACATCCACTT 1735
Db 2405 TCATCATGGACATGTTCAGGGACCCCAACACTTCAGTTGGTTCACCTCCACATCCACTT 2464
Qy 1736 TCAGGCTCAGGGGAACCAACCTCCCACTAGTGGGAACCTTACAATACACTTCTCTCCA 1795
Db 2465 TCAGGCTCAGGGGAACCAACCTCCCACTAGTGGGAACCTTACAATACACTTCTCTCCA 2524
Qy 1796 GGACTGACAGTGTCTCTCAGCCCAAGCCCACTAGTATATACCCCAAGCTGGGAAGCCAG 1855
Db 2525 GGACTGACAGTGTCTCTCAGCCCAAGCCCACTAGTATATACCCCAAGCTGGGAAGCCAG 2584
Qy 1856 GTCTACTGCTCCCAAGCAATTTGTGTACAGGTGCCACAGACACACAAGAATATCAG 1915
Db 2585 GTCTACTGCTCCCAAGCAATTTGTGTACAGGTGCCACAGACACACAAGAATATCAG 2644
Qy 1916 GAGCAGAAAGGATGGGAATGTGATGTTTAAAGAAATCCTTTGA 1962
Db 2645 GAGCAGAAAGGATGGGAATGTGATGTTTAAAGAAATCCTTTGA 2691

RESULT 3

LOCUS

DEFINITION

BC029658

Homo sapiens, Similar to endothelial and smooth muscle cell-derived

neurotrophin-like protein, clone MGC:30023 IMAGE:4431275, mRNA,

complete cds.

ACCESSION

BC029658

VERSION

BC029658.1

KEYWORDS

MGC.

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3858)

Strausberg, R.

Direct Submission

Submitted (06-MAY-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTF

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>Contact: amg@bcm.tmc.edu

Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,

Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,

Richards, S., Gibbs, R.A.

Clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 42 Row: c Column: 3.

FEATURES

source

1. 3858

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:30023 IMAGE:4431275"

/tissue_type="Prostate, adenocarcinoma."

/clone_lib="NIH_MGC_91"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

21..2252

/codon_start=1

CDS

BASE COUNT	1137 a	807 c	785 g	1129 t	
ORIGIN	E11"				
	/product="Similar to endothelial and smooth muscle				
	cell-derived neurofilin-like protein"				
	/protein_id="AA029658.1"				
	/db_xref="GI:2098615"				
	/translation="MPFLVLLVLLVLLDAGAAQDGGHVTLPGESGTLTSINYP				
	QTPVNSVGEWEIRKMGVRVKFGDFIEDSDCHFNLYRINYGVSRTLSIRGYC				
	GLGLONNIESKNGNLTILFMSGIHVSGRGLASVYIDKQDLITCLDTASNFLEPE				
	FSKYCAPGILLPFAEISGTPHGRDSSPLCMAGVHGVVNTLGGQISVIVISKGPY				
	YESSLANNTSVVGHLSISLFTPTKSGCTGLMGESGIVANTLQSVIEMDTHGQ				
	ENSWKPKARKLKPGPKAAFTDEYQWLQIDLNKCKKITGLTTCSTWVEHNVYSA				
	YRTLYSDDGQWTVYREPGEVDQIFQGNKQIDYVQVNRNFFLPILIAFTIRVETQMQ				
	KIAMKELGCCQFIPKGRPKLQPPBPNSNDLKNTPAPKIAKGAKPTQLOPR				
	KSNEFPAQEQTTPASDINTVTPNVTKDVALAALVLPVLVWLTLLILLVCANHN				
	KRNKKETIDLPYDWRAGFVLYWLSLACHNNEGWMKMQFLPAKADHDEETPVRI				
	SPQVNLSPREVTVTQDASEAQAFLGVGIVTGLHQSTFKPEEKREAGYADLDPIN				
	SPQEVYHAYAEPLTIGDSEATPIIMDSMGHPTTVGQSTSTFKATGNQPPPLVGT				
	VNTLLSRTDSCSAQAQYDTPKAGKPLGAPDELVYVQPQSTQEVSGAGRDGCDVFK				
	E11"				

BASE COUNT	1137 a	807 c	785 g	1129 t	ORIGIN
Query Match	93.9%	Score 1842.2	DB 9	Length 3858	
Best Local Similarity	97.4%	Pred. No. 0			
Matches 1899	Conservative	0	Mismatches 8	Indels 42	Gaps 1
Qy	56	GAGCCAGCAGGCAATAC	TGCTGCTGGGGTGC	AAATCAACCAATCAAT	TGAATCAA 115
Db	304	GAATGAATAGGCAATAC	TGCTGCTGGGGTGC	AAATCAACCAATCAAT	TGAATCAA 363
Qy	116	AAGCAATGAATCACAT	TGCTGCTAGTGGAA	TCCATGTTCTGGACG	CGGATTTT 175
Db	364	AAGCAATGAATCACAT	TGCTGCTAGTGGAA	TCCATGTTCTGGACG	CGGATTTT 423
Qy	176	TGGCCTCATACTCTG	TATAGATAAACAAG	ATCFAATTAATCTG	TTTGGACACTGCATCCA 235
Db	424	TGGCCTCATACTCTG	TATAGATAAACAAG	ATCFAATTAATCTG	TTTGGACACTGCATCCA 483
Qy	236	ATTTTGGAAACCTGAG	TTCAAGTACTGCCA	GTGGTGTGCTGCTT	CCTTTTGCTG 295
Db	484	ATTTTGGAAACCTGAG	TTCAAGTACTGCCA	GTGGTGTGCTGCTT	CCTTTTGCTG 543
Qy	296	AGATATCTGGAAACA	ATTCCTCATGGAT	ATAGAGATTCCTG	CCCATTTGTCATGGCTG 355
Db	544	AGATATCTGGAAACA	ATTCCTCATGGAT	ATAGAGATTCCTG	CCCATTTGTCATGGCTG 603
Qy	356	TGCATGCAAGGAGT	AGTGTCAAACAG	TTGGCGGCGCAAT	CAGTGTGTAATAGTAAG 415
Db	604	TGCATGCAAGGAGT	AGTGTCAAACAG	TTGGCGGCGCAAT	CAGTGTGTAATAGTAAG 663
Qy	416	GTATTCCTATATGAA	AGTTCTTTGGCT	TAACAGCTCACAT	CTGCTGCTGGACACTTAT 475
Db	664	GTATTCCTATATGAA	AGTTCTTTGGCT	TAACAGCTCACAT	CTGCTGCTGGACACTTAT 723
Qy	476	CTACAAGTCTTTTAC	ATTAAAGCAAGT	GTGTTATGGAAC	ACTGGGGATGGAGTCTG 535
Db	724	CTACAAGTCTTTTAC	ATTAAAGCAAGT	GTGTTATGGAAC	ACTGGGGATGGAGTCTG 783
Qy	536	GTGTGATCGCGGAT	CCTCAATACAGCA	TATCTGTGCTGG	ATGAGTACTGACCAACAG 595
Db	784	GTGTGATCGCGGAT	CCTCAATACAGCA	TATCTGTGCTGG	ATGAGTACTGACCAACAG 843
Qy	596	GGCAAGAAACAGT	TGGAAACCCAAA	AAAGCCAGGCTCA	AAAAAACCCTGGACCGCCTTGGG 655
Db	844	GGCAAGAAACAGT	TGGAAACCCAAA	AAAGCCAGGCTCA	AAAAAACCCTGGACCGCCTTGGG 903
Qy	656	CTGCTTTTGGCCACT	GATGAATACCACT	GGTTTACAAATAG	ATTGTAATAGGAAAAGAAAA 715
Db	904	CTGCTTTTGGCCACT	GATGAATACCACT	GGTTTACAAATAG	ATTGTAATAGGAAAAGAAAA 963
Qy	716	TAAACAGGCATTATA	ACCACCTGGATCC	ACCATTGGTGGAC	ACAAATTTACTATGTCTGCCT 775
Db	964	TAAACAGGCATTATA	ACCACCTGGATCC	ACCATTGGTGGAC	ACAAATTTACTATGTCTGCCT 1023

```

|||||
Db 2104 CAGCCAGGCCAGTAGTATACCCGAAAGCTGGGAAGCCAGGTCTACCTGCCAGCAGC 2163
|||||
QY 1874 AATTGTGTACCAAGTGGCCACAGAGCACACAAGAGTATCAGGACGAGGAAGGATGGGG 1933
|||||
Db 2164 AATTGTGTACCAAGTGGCCACAGAGCACACAAGAGTATCAGGACGAGGAAGGATGGGG 2223
|||||
QY 1934 AATGTGATGTTTTTAAGAAATCCCTTTGA 1962
|||||
Db 2224 AATGTGATGTTTTTAAGAAATCCCTTTGA 2252
|||||

RESULT 4
AF387549 2339 bp mRNA linear ROD 11-NOV-2001
LOCUS Rattus norvegicus endothelial and smooth muscle cell-derived
DEFINITION neuropeilin-like protein (Esdn) mRNA, complete cds.
ACCESSION AF387549
VERSION AF387549.1 GI:16902438
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 2339)
Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N.,
Matsumori,A., Sasayama,S., Honjo,T. and Tashiro,K.,
ESDN, a novel neuropeilin-like membrane protein cloned from vascular
cells with the longest secretory signal sequence among eukaryotes,
is up-regulated after vascular injury
J. Biol. Chem. 276 (36), 34105-34114 (2001)
21424027
PUBMED 11447234
REFERENCE
2 (bases 1 to 2339)
Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N.,
Sasayama,S., Matsumori,A., Honjo,T. and Tashiro,K.,
Direct Submission
Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto
University Graduate School of Medicine, Yoshida Konoe-cho,
Sakyo-ku, Kyoto 606-8501, Japan
FEATURES
source
1..2339
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1..2339
/gene="Esdn"
12..2321
/codon_start=1
/product="endothelial and smooth muscle cell-derived
neuropeilin-like protein"
/protein_id="P430180.1"
/db_xref="GI:16902439"
/translation="MASRAPLRAARSPPDGGRAAPAAATGRAPLPSAGWCPLPPGRNS
SSRPRLLLLLLPDAGAQKDCGGHVLGPESGTLTSINYPHTYPMNSTVCKWEIRV
KTERIRIKFGDFDIEDSDYKLVNLFKIFNGIVSRTEIGKYGLGQMNOISIESKGS
EITVLVMSIHASGRGFLASVVIDKODLITCLDTVSNFLEPEFSKPCACGLFPFAE
ISGTPHGRVSSPLCMAGIHAGVVDVVGQISVVISKGPYTESIANNVTSWVG
LSTSLFTKSGCGTGLGMSGVADPOITASSVLEWTDHMGQENSKPKERARKPG
PPWAAFDQHEQWLOIDLNEKKITGIVTGSTLTIEHNYVYSAYRVLSDDQKWTYV
REPAAQDKIFQGNKDHKDNFNLFP IIAFRVIRVNPVQWQKIAKVELLGCQFTL
KGRPLKTPPPRNSNLLKNTVHPKLGRAKFTQALQPSRNDLPLPAOTATPD
VKNTVTPSVTKDALAVLPLVLMALTTLLILVCAWHNRNKKAEQYDPLPHD
RAGWKVKQLLPKAKSVHEETPVRYNSVSHLSPREVTVLQADSREYAPLPGGI
VGLTHQSTFPEKGEKASYADLPYNAPVOEVYHAYAEPLPVTGPEYATPVMDSG
HSTASVGLPSTSTFTAGNQPPALVGTINTLLSRDSCSSQAQYDTPKGGKPAAPAE
ELTVQVQSTQEAAGCGRDEKFAKETL"
612 a 606 c 594 g 527 t

BASE COUNT 612 a 606 c 594 g 527 t
ORIGIN

Query Match 73.4%; Score 1440.2; DB 10; Length 2339;

```

```

Best Local Similarity 85.5%; Pred. No. 0;
Matches 1630; Conservative 0; Mismatches 268; Indels 9; Gaps 2;

QY 56 GAGCCAGCAGGCAAGCAATACTGTGGTCTGGGGTTGCAAAATGAACCAATCAATTAATCA 115
Db 424 GAACGGAATAAGCAAGTACTGTGGTCTGGGGTTTACAAAATGAATCAATCAATTAATCA 483
QY 116 AAGGCAATGAATCAATCAATGCTGTTCATGAGTGAATCCATGTTCTGGACGCGGATTTT 175
Db 484 AAGGCAATGAATCAATCAATGCTGTTCATGAGTGAATCCATGTTCTGGTCTGAGGATTTT 543
QY 176 TGGCCTCACTACTCTGTTATAGATAAACAAGATCAATTAATCTGTTTGGACACTGCATCCA 235
Db 544 TGGCTCTTACTCAGTATAGATAAACAAGATTAATCAATCAATCTGTTTGGATCTGATCTA 603
QY 236 ATTTTGGAACTGAGTTCAGTAACTAGTACTGCCACCTGGTGTCTGCTCTCTCTTTGCTG 295
Db 604 ATTTTGGAACTGAGTTCAGTAACTAGTACTGCCACCTGGTGTCTGCTCTCTCTTTGCTG 663
QY 296 AGATATCTGGAACAATTCCTCATGATATAGAGATTCCTCGCCATGTGTCATGGCTGGTG 355
Db 664 AATATCTGGAACGATTCCTCATGATATAGAGATTCCTCGCCGCTGTGTATGGCTGGAA 723
QY 356 TGCATGCAAGAGTAGTGTCAACACAGCTTGGCGGCCAAATCAGTGTGTAAATAGTAAAG 415
Db 724 TCCATGCAAGAGTAGTGTCAAGATGTGCTGGGTGGCCAAATCAGCGTTGTGATTAGCAAAG 783
QY 416 GTATTCCCTATTAATGAAAGTCTTGGCTTAACACGTCACATCTGTGGTGGGACACTTAT 475
Db 784 GCACCCCATATACGAAAGTCTTGGGCCAACAAATCTACCTCCATGGTGGGATACATAT 843
QY 476 CTACAAGTCTTTTACATTAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTG 535
Db 844 CTACAGTCTGTTTACATTAAGACAAGTGGTGTGCTATGGACTCTAGGATGGAGTCTG 903
QY 536 GTGTATCGCGGATCCTCAAAATACAGCATCATCTGCTGGTGGAGTGCAGTGCACACAG 595
Db 904 GTGTATCGCGGATCCTCAAGATACAGCATCATCTGCTGGTGGAGTGCAGTGCACACAG 963
QY 596 GCGAAGACAGTGGAAACCCAAACCCAGGCTGAAACACCTGGACCCCTGGG 655
Db 964 GCGAAGACAGTGGAAACCCAGGCTGAAACACCCAGGCTGAAACACCCGCGCTCCCTGG 1023
QY 656 CTGCTTTTGGCACTGATGAATACCACTGGTGTACAAATAGATTTGAATAAGAAAGAAA 715
Db 1024 CTGCTTTTGGCACTGATGAGCATCAGTGGCTGCAATTTGACCTTATTAAGGAAGAA 1083
QY 716 TAACAGGCAATTAACCACTGGATCCACCATGGTGGAGCACAAATTAATGCTGTGCTCT 775
Db 1084 TAACAGGCAATTAACCACTGGATCCACCATGGTGGAGCACAAATTAATGCTGTGCTCT 1143
QY 776 ACAGATCTGTACAGTGTATGGCAGAAATGGACTGTGTACAGAGAGCCCTGGTGTGG 835
Db 1144 ACAGATCTGTACAGTGTATGGCAGAAATGGACTGTGTACAGAGAGCCCTGGTGTGG 1203
QY 836 AGCAAGATAAGATATTTCAAGAAACAAAGATTTATCACCAGGATGTGCGTAAATCTTT 895
Db 1204 CTCAGGACAGATATTTCAAGAAACAAAGATTTATCACAAGGATGTGCTTAATCTTT 1263
QY 896 TCCACCAATTAATGACCTTTTATTTAGAGTGAATCCTACCCCAATGGCAGACGAAATTTG 955
Db 1264 TCCACCAATTAATGACCTTTTATTTAGAGTGAATCCTACCCCAATGGCAGACGAAATTTG 1323
QY 956 CCATCAAAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1015
Db 1324 CCATCAAAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1383
QY 1016 CTCACCTCCCACTCCTCGGAACAGCAATTAAGCTCAAAACACTACAGCCCTCCCAAAA 1075
Db 1384 CTCACCTCCCACTCCTCGGAACAGCAATTAAGCTCAAAACACTACAGTTCATCCCAAC 1443
QY 1076 TAGCAAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1135

```


Db 966 AATATCTGGACAAATTCCTCATGATACAGAGATTCTTCCACCATTTGTGTATGGCTGGAA 1025
QY 356 TGCATGAGAGTAGTGTCAAAACACGTTGGGGGCGCCAAATCAGTGTGTGTAATAGTAAG 415
Db 1026 TCCATGAGAGTAGTGTCAAAACGCTGGGGTGGCCAAATCAGCAATTTGATTAGCAAG 1085
QY 416 GTATTCCTTATTATGAAGTCTTTTGGCTAACACGTCACATCTGTGGTGGGACACTTAT 475
Db 1086 GGACCCCATATTATGAAGTCTTTTGGCCCAACCAATTCACATTCACGGTGGGATACATAT 1145
QY 476 CTCAAGTCTTTTACATTTAAAGCAAGTGGATGTTATGGAACACTGGGATGGAGTCTG 535
Db 1146 CTGCAAGTCTTTTACATTTAAGACAGTGGTTCCTATGGACTCTGGGATGGAGTCTG 1205
QY 536 GTGTGATCGGGATCTCAATTAATACACGATCATCTGTGCTGGAGTGGACTGACACACAG 595
Db 1206 GTGTGATTCGGATCCCGAGATACACGATCTGTCTGCACTGGAGTGGACTGACACATGG 1265
QY 596 GGCAAGAGAACAGTTGGAAACCCCAAAAGCCAGGCTGAAAGAACCTGGAGCCGCTTGGG 655
Db 1266 GGCAGGAGACAGCTGACACGGGAGAGGCCAGGCTGAGAAACCCGGGCTTCCCTGGG 1325
QY 656 CTGCTTTTGGCACTGATGAATACACAGTGGTTACAAATAGATTTGAATPAAGGAAAGAAA 715
Db 1326 CTGCTTTTGGCACTGATGAGCATCAGTGGCTGCAGATAGACCTTAACAAGGAGAAAGA 1385
QY 716 TAACAGGCAATATACACACTGGATCCACCATGGTGGAGCAATTAATCTATGTCTGCT 775
Db 1386 TAACAGGATTCGTAACCACTGGGTCTACCATGATAGAACAGATTACTATGTCTGCT 1445
QY 776 ACAGAAATCCTGTACAGTGTATGGGAGAAATGGACTGTGTACAGAGAGCCCTGGTGTG 835
Db 1446 ACAGATCCTGTACAGTGTATGGGAGAGATGGACTGTGTACAGAGAACCTGGTGTG 1505
QY 836 AGCAAGATAGATATTTCAAGGAACAAAGATATACACAGAGTGTGCGTAATACATTTT 895
Db 1506 ACCAGGACAGATATTTCAAGGAACAAAGATATACAGAGGATGTTGCTTAATACATTT 1565
QY 896 TGGCACCATAATTGACAGTTTATTAGAGTGAATCTTACCCATGGCAGCAGAAATTTG 955
Db 1566 TGGCACCATAATTGACAGTTTATTAGAGTGAATCTTACCCATGGCAGCAGAAATTTG 1625
QY 956 CCATGAAATGGAGCTGCTCGGATGTCTAGTTTATCTTAAAGTGTCTCTCCCAAACTTA 1015
Db 1626 CCATGAAATGGAGCTGCTCGGATGTCTAGTTTATCTTAAAGTGTCTCTCCCAAGCTTA 1685
QY 1016 CTCAACTCCACCTCTCGGAGACAGCAATGACCTCAAAACACTACAGCCCTCCCAAAA 1075
Db 1686 -----CTCCACCTCTCTCGGAGCAGCAATGACCTCAAAATACTACAGCTCTGCCCAAAC 1739
QY 1076 TAGCCAAAGTCTGCCCCCAAAATTTACGCAACCACTACACCTCGCAGTAGCAATGAAT 1135
Db 1740 TAGTAAAGTGTGCGCCCTTAAATTTACTCAAGTGTCTTCAACCTCGAAGTAGGAATGAAC 1799
QY 1136 TTCTGTGCACAGACAGAAACAACTGCCAGTCTCTGATATCAGAAATACCTACCGTAACCT 1195
Db 1800 TTCTGTGTGAGCCGCGGAGACACTACACTCTCTGATATAAAACACAGCTGTAACT 1859
QY 1196 CAATGTAACCAAGATGTAGCGCTGGCTGCAGTCTTTGTCCCTGTGCTGTGATGTGCTC 1255
Db 1860 CCAGTGTAAACCAAGATGTGCAGCTGGCTGCCGTTCTGTGCTCCCTGTGCTGTGCTGCC 1919
QY 1256 TCACACTCTCTATCTCTATATTAGTGTGCTTGGGACCGGCGAGTGTGGTGAAGGATGAAC 1315
Db 1920 TCACCACTACTCTCTCTATCTAGTGTGCTTGGCACTGGAGAAACAGGAAAGAAA 1979
QY 1316 CTGAAGGACCTATGACTTACCTTACTGGGACCGGCGAGTGTGGTGAAGGATGAAC 1375
Db 1980 CTGAAGGCGCTATGATTTACCCCACTGGGATCGGGAGTGTGGTGAAGGATGAAC 2039
QY 1376 AGTTCTCTCTGCAAAAGAGTGGACCATGAGGAACCCAGTTCGCTATAGACAGCAGCG 1435
Db 2040 AGCTTCTCTGCAAGTGGTGGACACAGGAGAGCCAGTGGCTGTACGACACTAGT 2099

QY 1436 AAGTTAATCACTGAGTCCAAAGAGAAAGTACCACAGTGTGTCAGGCTGACTTCCAGAGT 1495
Db 2100 AAGTCACTGCTGAGTGCAGGGAAGTACCACAGTGTGTCAGGCGGAGCTCTCAGAAAT 2159
QY 1496 ATGCTAGCCACTGGTAGGAGGAATTTGTGTACACTTCATCAAGACTCTACCTTTTAAAC 1555
Db 2160 ATGCACAGCCCCCTCGTGGGAGGAATTTGTGGCACACTCCATCAGAGATCCACCTTTAAAC 2219
QY 1556 CAGAAGAAAGAAAGAACAGAGGCTATGCAGACCTTAGATCTTTTACAACTCACCAGGCGAG 1615
Db 2220 CTGAGGAAGGAAGGAAGCAGGCTATGCAGACCTCGATCTTACAACTCTCCTCAATGCAG 2279
QY 1616 AAGTTTATCATGCTATGCTGAACCACTCCCAATTTAGGGGCTGAGTATGCACACCCAA 1675
Db 2280 AAGTGTACCAAGGCTATGCTGAACCACTGCGCTTAACGGGCTGAGTACGACCAACCCGA 2339
QY 1676 TCATCATGGACATGTCTAGGGCAGCCCAACAACTTCAGTTGGTGGTGGTGGTGGTGGTGG 1735
Db 2340 TCGTATGACATGTCTAGGGCAGCCCAACCACTTCAGTTGGTGGTGGTGGTGGTGGTGG 1795
QY 1736 TCAAGGCTACGGGAACCAACCTCCCACTAGTGGGAATTAACATACACTTCTCTCCA 1795
Db 2400 TCAAACTGCGAGGACCCAGCCTCACGCTTTAGTGGAACTTACAACACTTCTCTCCA 2459
QY 1796 GGACTGACACTGCTCTCAGCCAGCCAGCTATGATACCCGAAAGCTGGGAAGCCAG 1855
Db 2460 GGACTGACACTGCTCTCAGCCAGCCAGCTATGATACACCCCAAGGTTGGGAAGTCAG 2519
QY 1856 GTCTACCTGCCAGCAGCAATTTGGTGTACAGGTCGCCACAGCAGCAGCAAGATATCAG 1915
Db 2520 CT---GCTACCCAGAGAACTGGTATACAGGTCGCCACAGCAGCCAGGAGTATCAG 2576
QY 1916 GAGCAGAAAGGATGGGAATGTGATGTTTTAAAGAAATCCTTTGA 1962
Db 2577 GAGCAGAAAGGATGAGAAGTTGATGCTTTTAAAGAAATCCTTTGA 2623
RESULT 6
AK056350
LOCUS
DEFINITION
to MLTK FAT GLOBULE-EGF FACTOR 8 PRECURSOR.
ACCESSION
AK056350
VERSION
AK056350.1 GI:16551729
KEYWORDS
oligo capping; fls (full insert sequence).
SOURCE
Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
clone_lib:NT2R12 clone:NT2R12008598.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujil,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahara,K., Masubo,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2020)
Isogai,T., Otsuki,T. and Sugiyama,T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan

Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

```

FEATURES             Location/Qualifiers
     source            1..2020
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="NT2R12008598"
                        /cell_line="NT2"
                        /cell_type="teratocarcinoma"
                        /clone_lib="NT2R12"
                        /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
                        precursor cells treated 2-weeks mitotic inhibitor after
                        5-weeks retinoic acid (RA) induction.-majority NT2 neuron"
BASE COUNT          494 a 519 c 530 g 477 t
ORIGIN
Query Match          52.7%; Score 1230.2; DB 9; Length 2020;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1235; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 56 GAGCCAGCAAGCAACTACTGTGGTCTGGGTTGCAAAATGAACCATTCGAATCAAA 115
DB 778 GAACAGTAAGCAAACTACTGTGGTCTGGGTTGCAAAATGAACCATTCGAATCAAA 837
QY 116 AAGCAATGAATACACATTCCTGTTCTATGAGTGAATCCATGTTCTCGACGCGGATTT 175
DB 838 AAGCAATGAATACACATTCCTGTTCTATGAGTGAATCCATGTTCTCGACGCGGATTT 897
QY 176 TGGCCTCATACTCTGTATAGATAAACAAGATCTAATTAATTTGTTGGACACTGCATCCA 235
DB 898 TGGCCTCATACTCTGTATAGATAAACAAGATCTAATTAATTTGTTGGACACTGCATCCA 957
QY 236 ATTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTCTCTCTTTGCTG 295
DB 958 ATTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTCTCTCTTTGCTG 1017
QY 296 AGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTGCCATTTGTCATGCTGGTG 355
DB 1018 AGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTGCCATTTGTCATGCTGGTG 1077
QY 356 TGCATCAGAGTAGTCTCAACACGTTGGCGGCCAAATCAGTGTGTAAATAGTAAG 415
DB 1078 TGCATCAGAGTAGTCTCAACACGTTGGCGGCCAAATCAGTGTGTAAATAGTAAG 1137
QY 416 GTATTCCTTATATGAAGTCTTTGGCTAACACGTCACATCTGTGTGGGACACTTAT 475
DB 1138 GTATTCCTTATATGAAGTCTTTGGCTAACACGTCACATCTGTGTGGGACACTTAT 1197
QY 476 CTACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTG 535
DB 1198 CTACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTG 1257
QY 536 GTGTGATCGGGATCTCAATAACAGCATCATCTGTCTGGAGTGGACTGACCAACAG 595
DB 1258 GTGTGATCGGGATCTCAATAACAGCATCATCTGTCTGGAGTGGACTGACCAACAG 1317
QY 596 GGCAAGAGAACAGTTGGAACCCAAACAGCGCTGAAACACCTGGACCGCTTGGG 655
DB 1318 GGCAAGAGAACAGTTGGAACCCAAACAGCGCTGAAACACCTGGACCGCTTGGG 1377
QY 656 CTGCTTTGGCACTGATGAATACCAAGTGGTTACAAATAGATTTTGAATAAGGAAAGAAA 715
DB 1378 CTGCTTTGGCACTGATGAATACCAAGTGGTTACAAATAGATTTTGAATAAGGAAAGAAA 1437
QY 716 TAACAGGCATTAACCACTGGATCCACCATGTTGGAGACCAATTAATTAATGTGTGCTGCT 775
DB 1438 TAACAGGCATTAACCACTGGATCCACCATGTTGGAGACCAATTAATTAATGTGTGCTGCT 1497
QY 776 ACAGAACTCTGTACAGTATGATGGGAGAAATGGACTGTGTACAGAGAGCTGTGGTGG 835
DB 1498 ACAGAACTCTGTGAGTATGATGGGAGAAATGGACTGTGTACAGAGAGCTGTGGTGG 1557

```

```

QY 836 AGCAAGATAAGATATTTCAAGGAACAAAGATTATCACCAGGATGTCGCTAATAACTTTT 895
DB 1558 AGCAAGATAAGATATTTCAAGGAACAAAGATTATCACCAGGATGTCGCTAATAACTTTT 1617
QY 896 TGCACCAATATTTTGCACGTTTATTAGAGTGAATCTTACCCTATGCGAGCAGAAAAATTG 955
DB 1618 TGCACCAATATTTTGCACGTTTATTAGAGTGAATCTTACCCTATGCGAGCAGAAAAATTG 1677
QY 956 CCATGAAAATGGAGCTGCTCGGATGTCTGAGTTTATTTCTAAAGGTCGCTCCCAAACTTA 1015
DB 1678 CCATGAAAATGGAGCTGCTCGGATGTCTGAGTTTATTTCTAAAGGTCGCTCCCAAACTTA 1737
QY 1016 CTCAACTCCACCTCTCGGAACAGCAATGACCTCAAAACACTACAGCCCTCCCAAAA 1075
DB 1738 CTCAACTCCACCTCTCGGAACAGCAATGACCTCAAAACACTACAGCCCTCCCAAAA 1797
QY 1076 TAGCCAAAGTCTGCGCCCAAAATTTAGCAACCACTACACCTCGCAGTAGCAATGAAT 1135
DB 1798 TAGCCAAAGTCTGCGCCCAAAATTTAGCAACCACTACACCTCGCAGTAGCAATGAAT 1857
QY 1136 TTCCTGCACAGACAGAAACAACACTGCCAGTCTCTGATATCAGAAATACCTACCGTAATC 1195
DB 1858 TTCCTGCACAGACAGAAACAACACTGCCAGTCTCTGATATCAGAAATACCTACCGTAATC 1917
QY 1196 CAAATGTAACCAAGATGTAGCGCTGGCTGCAGTCTTGTGCTCCCTGCTGCTCATGCTCC 1255
DB 1918 CAAATGTAACCAAGATGTAGCGCTGGCTGCAGTCTTGTGCTCCCTGCTGCTCATGCTCC 1977
QY 1256 TCAGTACTCTCATCTCATATATAGTGTGCTTGGCACTGGAG 1298
DB 1978 TCAGTACTCTCATCTCATATATAGTGTGCTTGGCACTGGAG 2020

RESULT 7
AX405795
LOCUS AX405795 1907 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 210 from Patent WO0222660.
ACCESSION AX405795
VERSION AX405795.1 GI:21439029
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,
Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.
TITLE Novel nucleic acids and polypeptides
JOURNAL Patent: WO 0222660-A 210 21-MAR-2002;
HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source            1..1907
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        101..1258
                        /note="unnamed protein product"
                        /codon_start=1
                        /protein_id="CAD34902.1"
                        /db_xref="GI:21439030"
     translation="MPLFLLLLVLLLLLEDAGAQQDGGCHTVLGPESGTLTSINYP
     QTPNSTVCEWEIRVKMERVRIKFGDFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYC
     GLQNMNHSIESKNEITLLFMGSIHVSGRGLASVYIDKDLITCLDTPASNFLPE
     FSKVPAGCLLPFAEISGTIPHYGRDSSPLCMAGVHGVVNTLGGIISVVISKGIPI
     YESSLANNTSVVGHLSLTSLFTKSCYCTGLMGESGDRGSSNNTITVLEWIDHTGQ
     ENSWKPKKSQREKNTALGAFATDEIOWLQIDLKKEKKTGIITGOSTVSTVMTMCLP
     TESCVMGMRGLCTESLVVEQKIFGKNRIITRMVRNNFLPIIARLLE"
BASE COUNT          548 a 449 c 406 g 504 t
ORIGIN
Query Match          59.1%; Score 1159.2; DB 6; Length 1907;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1286; Conservative 0; Mismatches 8; Indels 14; Gaps 11;

```

QY	56	GAGCCGAGCAAGCAAAATACTGTGTCTGGGGTTGCAAAATGAACCAATCAATTGAATCAA	115
Db	384	GAATCGAAATAGGCAAAATACTGTGTCTGGGGTTGCAAAATGAACCAATCAATTGAATCAA	443
QY	116	AAGGCAATGAATACATTCGTTTCATGAGTGGAAATCCATGTTTCTGGAGCGGGATTTT	175
Db	444	AAGGCAATGAATACATTCGTTTCATGAGTGGAAATCCATGTTTCTGGAGCGGGATTTT	503
QY	176	TGGCCTACTACTCTGTTATAGATAAACAGATCTAATTAATTTTGGACACTGCATCCA	235
Db	504	TGGCCTACTACTCTGTTATAGATAAACAGATCTAATTAATTTTGGACACTGCATCCA	563
QY	236	ATTTTTTGGAACTGAGTTTCAGTAAGTACTGCCAGCTGGTTGCTGCTTCTCTTTTGGCTG	295
Db	564	ATTTTTTGGAACTGAGTTTCAGTAAGTACTGCCAGCTGGTTGCTGCTTCTCTTTTGGCTG	623
QY	296	AGATATCTGGAACAATTCCTCATGGATATAGAGATTCTCGCCCATTTGTGATGGCTGGTG	355
Db	624	AGATATCTGGAACAATTCCTCATGGATATAGAGATTCTCGCCCATTTGTGATGGCTGGTG	683
QY	356	TGCATGACGAGTAGTGTCAAACACGTTGGGGGGCCAAATCAGTGTGTAATTAAGTAAG	415
Db	684	TGCATGACGAGTAGTGTCAAACACGTTGGGGGGCCAAATCAGTGTGTAATTAAGTAAG	743
QY	416	GTATTCCTTATATGAAGTTCCTTTGGCTTAACAAGTACACATCTGTGGTGGGACACTTAT	475
Db	744	GTATTCCTTATATGAAGTTCCTTTGGCTTAACAAGTACACATCTGTGGTGGGACACTTAT	803
QY	476	CTACAAGTCTTTTACATTTTAAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTG	535
Db	804	CTACAAGTCTTTTACATTTTAAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTG	863
QY	536	GT-GTATCGCGGATCCTCAAATAACAGCATCATCTGTGCTGGAGTGGACTGCACACACA	594
Db	864	GTGATGATCGCGATCCTCAAATAACAGCATCA-CTGTGCTGGAGTGGACTGCACACACA	922
QY	595	GGCAGAGNACAGTTGGAAACCC-AAAAAGCCAGGCTGAAAAAACCCTGGACGCCCTTG	653
Db	923	GGCAGAGNACAGTTGGAAACCCAAAAAGCCAGGCTGAAAAAACCCTGGACGCCCTTG	982
QY	654	GGGTGCTTTTGGCACTGATCAATACAGTGGTTCACAAATAGATTGTAATAGGAAAAAGAA	713
Db	983	GG-TGCTTTTGGCACTGATCAATACAGTGGTTCACAAATAGATTGTAATAGGAAAAAGAA	1041
QY	714	AATAACAGGCAATTATAACCACTGGATCCACCATGGTGGAGCACAAATTAATGTGTCTG	773
Db	1042	AATAACAGGCAATTATAACCACTGGATCCACCATGGT-GAGCACAAATTAATGTGTCTG	1100
QY	774	CTACAGATCCTGTACAGTGTATGGGCGAGAAATGGACTGTGTACAGAGAGCCTGCT-G	832
Db	1101	CTACAGATCCTGTACAGTGTATGGGCGAGAAATGGACTGTGTACAGAGAGCCTGCTG	1160
QY	833	TGGAGCAAGATAAGATATTTCAGGAAACAAA--GATTATCACCAGGAT-GTGGCTAATA	889
Db	1161	TGGAGCAAGATAAGATATTTCAGGAAACAAAAGAAATTAACACAGGATGTGGCTAATA	1220
QY	890	ACTTTTGGCACCAATTTATGACGCTTTTATTAGAGTGAATCTTACCCAAATGGCAGAGA	949
Db	1221	ACTTTTGGCACCAATTTATGACG-TTTTATTAGAGTGAATCTTACCCAAATGGCAGAGA	1279
QY	950	AAATTGGCATGAAATGGAGCTGCTCGGATGTCAGTTTATCTTAAAGGTGCTCTCTCAA	1009
Db	1280	AAATTGGCATGAAATGGAGCTGCTCGGATGTCAGTTTATCTTAAAGGTGCTCTCTCAA	1339
QY	1010	AACCTACTCAACCTCCACCTCTCGGAAACAGCAATTAAGCTCAAAACACTACAGCCCTC	1069
Db	1340	AACCTACTCAACCTCCACCTCTCGGAAACAGCAATTAAGCTC-AAAAACACTACAGCCCTC	1398
QY	1070	CAAAAA--TAGCCAAAGTCTGTCGCCCAAAAATTTACGCAACCACTACAACCTGCGAGTA	1126
Db	1399	CAAAAAATTTAGCAAAAGTCTGTCGCCCAAAAATTTACGCAACCACTACAACCTGCGAGTA	1458

QY	1127	GCATGAATTTCCCTGCACAGACAGAAACAAACAACTCCAGTCCCTGATATCAGAATACTA	1186
Db	1459	GCAATGAATTTCCCTGCACAGACAGAAACAAACAACTCCAGTCCCTGATATCAGAATACTA	1518
QY	1187	CCGTAACATCCAAATGTAAACAAAGATGTAGCGCTGGCTGCAGTCTTGTCTCCTCTGCTGG	1246
Db	1519	CCGTAACATCCAAATGTAAACAAAGATGTAGCGCTGGCTGCAGTCTTGTCTCCTCTGCTGG	1578
QY	1247	TCATGGCTCCTCACACTCTCTCATTTCTCATATTAGTGTGTGTGGCACTGGAGAAACAGAA	1306
Db	1579	TCATGGCTCCTCACACTCTCTCATTTCTCATATTAGTGTGTGTGGCACTGGAGAAACAGAA	1638
QY	1307	AGAAAAAACTGAAGGCACCTATACCTTACTCTACTGGACCGGCGCAGGT	1356
Db	1639	AGAAAAAACTGAAGGCACCTATACCTTACTCTACTGGACCGGCGCAGGT	1688
RESULT 8			
HUMCUB1			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens cell line L121 cDNA to mRNA, clone cub-1.			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
Submitted (08-Apr-1994) to DDBJ by:			
Tatsuhiko Shibata			
National Cancer Center Reserch Institute			
1-1 Tsukiji 5 chome			
Chuo, Tokyo 104			
Japan			
Phone: 03-3542-2511 x4208			
Fax: 03-3248-2737.			
Location/Qualifiers			
1. .1388			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/cell_line="Li21"			
<1..1097			
/codon_start=3			
/product="unknown"			
/protein_id="BAAL18909.1"			
/db_xref="GI:704441"			
/translation="DGGCHTVLGPESGTLTSINYPOTYPNSTVCWEIRVKMGERVRI			
KGFDFIEDSDSCHFNILRYINGLGVSRTEIGKCGLGLOLNHNSIESKGNIEITLLFNS			
GIHVSYGGRTFASVVIDKORSNYLFHGCIOFWNLSSVSTAPACGLLPFVGHILQVFP			
GHRDSSPLCLMGHAGVAVSYNTLGGQISVVISKGPVYESSLIANNVTISVGHILQVFP			
TFKTSYGYCTGLMGESGDRSSNNSTVLEWTDHTGQENSWKPKKSOAKETWTALGAF			
ATDYOLOWIDLNKEKKTIGTITGTSTWVSTITWCLPTESCTVMGMRNGLCTSLSVVE			
QDKTFQGNKRITRMVRNFPPIIARLE"			
BASE COUNT			
432 a 289 c 302 g .364 t 1 others			
ORIGIN			
Query Match			
Best Local Similarity 47.5%; Score 932.8; DB 9; Length 1388;			
Matches 1136; Conservative 0; Mismatches 18; Indels 23; Gaps 17;			
QY	56	GAGCCAGCAGGCAAACTGTGCTGTGGGTGTCACAAATGAACATTCAATTAATCAAA	115

Db 217 GAAGTGAATAGGCAATTAAGTCTGGTCTGGGTTGCAAAATCAACCAATCAATGAATCAA 276
QY 116 AAGCAATGAATACACATCTGTTTCATGAGTGAATCCATGTTCTCGGACGCGGATTTT 175
Db 277 AAGCAATGAATACACATCTGTTTCATGAGTGAATCCATGTTCTCGGACGCGGATTTT 336
QY 176 TGGCCTCACTCTGTTTATAGATAAAC - AAGATCTAATTAATCTGTTTGGACATGCATCC 234
Db 337 TGGCCTCACTCTGTTTATAGATAAACAAAGATCTAATTAATCTGTTTGGACATGCATCC 396
QY 235 AA-TTTTTCGAACCTGAGTCTAGTAGTACTG - CCGAGCTGTTGCTGCTCTCTTTT 292
Db 397 AATTTTTTGAACCTGAGTCTAGTAGTACTGCCCCAGCTGTTGCTGCTCTCTTTT 456
QY 293 CTGAGATATCTGAACAATTCCTCATGATATAGAGATTCCTGCCATTTGTGCATGGCTG 352
Db 457 CTGAGATATCTGAACAATTCCTCATGATATAGAGATTCCTGCCATTTGTGCATGGCTG 516
QY 353 GTGTGATGAGAGTAGTCTCAACACGTTGGCGGCCAAATCAGTGTGTGAATTAGTA 412
Db 517 GTGTGATGAGAGTAGTCTCAACACGTTGGCGGCCAAATCAGTGTGTGAATTAGTA 576
QY 413 AAGTATTCCTTATATGAAGTCTTTTGGCTAACACGTCACATCTGTGTGGGACACT 472
Db 577 AAGTATTCCTTATATGAAGTCTTTTGGCTAACACGTCACATCTGTGTGGGACACT 636
QY 473 T-ATCTACAAGTC-TTTTTACATTTTAAAGCAAGTGTGATTTATGGAACACTGGGGATGG 529
Db 637 TGATCTACAAGTCTTTTTTACATTTTAAAGCAAGTGTGATTTATGGAACACTGGGNATGG 696
QY 530 AGTCTGGT-CTGATCGGGATCTCTCAATTAACAGCATCATCTGTGTGGAGTGGACATGC 588
Db 697 AGTCTGGTGTGTGTCGGGATCTCTCAATTAACAGCATCA-CTGTGTGTGGAGTGGACATGC 755
QY 589 CACAGGGGCAAGACAGTGTGGAACCC - AAAAAGCCAGGCTGAAACACCTCGACC 647
Db 756 CACAGGGGCAAGACAGTGTGGAACCC - AAAAAGCCAGGCTGAAACACCTCGACC 815
QY 648 GCCTTGGCTGCTTTTGGCACTGATGAATACCACTGTTTCAATAGATTTGAATAAGGA 707
Db 816 GCCTTGGG-TGCTTTTGGCACTGATGAATACCACTGTTTCAATAGATTTGAATAAGGA 874
QY 708 AAGAAATTAACAGGCAATTAACCACTGATCCACCATGTTGGAGACAAATTAATATGT 767
Db 875 AAGAAATTAACAGGCAATTAACCACTGATCCACCATGTTGGAGACAAATTAATATGT 933
QY 768 GTCTGCTCAGCAATCTGTACAGTGTGATGGCAGAAATGGACTGTGTACAGAGACC 827
Db 934 GTCTGCTCAGCAATCTGTACAGTGTGATGGCAGAAATGGACTGTGTACAGAGACC 993
QY 828 TGGT-GTGGAGCAAGATAGATATTTCAAGGAAACAAA - GATTATCACCAGGAT-GTGC 883
Db 994 TGGTGCTGGAGCAAGATAGATATTTCAAGGAAACAAAATTAATCACCAGGATGTGC 1053
QY 884 GTAATACTTTTGGCCACCAATTAATTCAGCTTTTATAGAGTGAATTCCTACCAATGGC 943
Db 1054 GTAATACTTTTGGCCACCAATTAATTCAGC - TTTATTAAGTGAATTCCTACCAATGGC 1112
QY 944 AGCAGAAATTCATGAAATGAGTGTGCTCGGATGCTGATGTTTCCCTAAAGGTGCT 1003
Db 1113 AGCAGAAATTCATGAAATGAGTGTGCTCGGATGCTGATGTTTCCCTAAAGGTGCT 1172
QY 1004 CTCAAAACCTTACTCAACCTCCACTCTCGGACAGCAATGACTCAAAAACACTACAG 1063
Db 1173 CTCAAAACCTTACTCAACCTCCACTCTCGGACAGCAATGACTCAAAAACACTACAG 1231
QY 1064 CCCCCTCAAAA - -TAGCCAAAGGTGCTGCCCAAAATTTACGCAACCACTACAACTC 1120
Db 1232 CCCCCTCAAAAATTTAGCCAAAGGTGCTGCCCAAAATTTACGCAACCACTACAACTC 1291
QY 1121 GCAGTAGCAATGATTTCTTCGACAGACAGAAACAACTGCGAGTCTCGATATCAGAA 1180
Db 1292 GCAGTAGCAATGATTTCTTCGACAGACAGAAACAACTGCGAGTCTCGATATCAGAA 1351

QY 1181 AT---ACTACCGTAACCTCAATGTAACTAAACCAAGATGT 1214
Db 1352 ATTACCTACCGTAACCTCAATGTAACTAAACCAAGATGT 1388

RESULT 9

AC013497

LOCUS

DEFINITION

AC013497

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC013497 192247 bp DNA linear HTG 26-MAR-2000
Homo sapiens chromosome 3 clone RP11-319J24 map 3, WORKING DRAFT
SEQUENCE, 11 unordered pieces.

AC013497 GI:7329299
HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 192247)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 3, clone RP11-319J24
Unpublished

2 (bases 1 to 192247)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Strange-Romann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (13-NOV-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6692347.

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L1852
Center clone name: 319-J-24
----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182870 bases at least Q40
Consensus quality: 187032 bases at least Q30
Consensus quality: 188925 bases at least Q20

Insert size: 194000; agarose-fp
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
consists of 11 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

----- 472: contig of 472 bp in length
1

```

* 473 572: gap of 100 bp
* 573 3945: contig of 3373 bp in length
* 3946 4045: gap of 100 bp
* 4046 11056: contig of 7011 bp in length
* 11057 11156: gap of 100 bp
* 11157 17728: contig of 6572 bp in length
* 17729 17828: gap of 100 bp
* 17829 30556: contig of 12728 bp in length
* 30557 30656: gap of 100 bp
* 30657 49294: contig of 18638 bp in length
* 49295 49394: gap of 100 bp
* 49395 64509: contig of 15115 bp in length
* 64510 64609: gap of 100 bp
* 64610 83042: contig of 18433 bp in length
* 83043 83142: gap of 100 bp
* 83143 112373: contig of 29231 bp in length
* 112374 112473: gap of 100 bp
* 112474 145663: contig of 33190 bp in length
* 145664 145763: gap of 100 bp
* 145764 192247: contig of 46484 bp in length.

```

FEATURES

source

```

1. .192247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone_lib="RPC1-11 Human Male BAC"
/clone="RP11-319J24"
1. .472
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
573. .3945
/note="assembly_fragment"
4046. .11056
/note="assembly_fragment"
11157. .17728
/note="assembly_fragment"
17829. .30556
/note="assembly_fragment"
30657. .49294
/note="assembly_fragment"
49395. .64509
/note="assembly_fragment"
64610. .83042
/note="assembly_fragment"
clone_end:17
vector_side:left"
83143. .112373
/note="assembly_fragment"
112474. .145663
/note="assembly_fragment"
145764. .192247
/note="assembly_fragment"

```

BASE COUNT 58524 a 38157 c 36840 g 57721 t 1005 others

```

Query Match 24.1%; Score 473.8; DB 2: Length 192247;
Best Local Similarity 99.8%; Pred. No. 3.4e-131;
Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1486 TCTGCAGATGATGCTCAGCCACTGGTAGGAGGAATGTTGGTACACTTCATCAAGATCT 1545
|||
Db 164971 TCCACAGATGATGCTCAGCCACTGGTAGGAGGAATGTTGGTACACTTCATCAAGATCT 15030
|||
QY 1546 ACCTTTAAACAGAGAGGAAAGAGGAGGCTATGCAGACCTAGATCCTTACAACTCA 1605
|||
Db 165031 ACCTTTAAACAGAGAGGAAAGAGGAGGCTATGCAGACCTAGATCCTTACAACTCA 165090
|||
QY 1606 CCAGGGCAGAGGATTTATCATGCTCTATGCTGACCACTCCCAATTACGGGCGCTAGAT 1665
|||
Db 165091 CCAGGGCAGAGGATTTATCATGCTCTATGCTGACCACTCCCAATTACGGGCGCTAGAT 165150
|||

```

```

QY 1666 GCAACCCCAATCATGACATGTCTCAGGCGACCCCAACACTTCTAGTTGGTGCAGCCCTCC 1725
|||
Db 165151 GCAACCCCAATCATGACATGTCTCAGGCGACCCCAACACTTCTAGTTGGTGCAGCCCTCC 165210
|||
QY 1726 ACATCCACACTTCAAGGCTAGGGGAAACCAACCTCCGCCACTAGTGGGAACCTTACAATACA 1785
|||
Db 165211 ACATCCACACTTCAAGGCTAGGGGAAACCAACCTCCGCCACTAGTGGGAACCTTACAATACA 165270
|||
QY 1786 CTCTCTCCAGGACTGACAGCTGCTCCTCAGCCAGCCAGCCAGCTATGATATACCCGAAAGCT 1845
|||
Db 165271 CTCTCTCCAGGACTGACAGCTGCTCCTCAGCCAGCCAGCTATGATATACCCGAAAGCT 165330
|||
QY 1846 GGAAGCCAGGTTACTGCTGCCAGACGAATGGTGTACCAGGTGCCACAGAGACACAA 1905
|||
Db 165331 GGAAGCCAGGTTACTGCTGCCAGACGAATGGTGTACCAGGTGCCACAGAGACACAA 165390
|||
QY 1906 GAAGTATCAGGACGAGGAGGATGGGAATGTGATGATTTTAAAGAAATCCTTTGA 1962
|||
Db 165391 GAAGTATCAGGACGAGGAGGATGGGAATGTGATGATTTTAAAGAAATCCTTTGA 165447
|||

RESULT 10
AC106728 193623 bp DNA linear PRI 25-MAY-2002
Homo sapiens 3 BAC RP11-319J24 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.
AC106728
AC106728.6 GI:21206081
HTG.
Homo sapiens
Homo sapiens
Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
1 (bases 1 to 193623)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Huijck,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseghed,H., Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenwo,S., Oguh,M., Okwodu,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M.,
Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E.,
Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Vallalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wlarczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,

```



```

repeat_region      /rpt_family="LIPB3"
                    25951..26456
repeat_region      /rpt_family="MER75"
                    26457..27565

```

```

Query Match      24.1%; Score 473.8; DB 9; Length 193623;
Best Local Similarity 99.6%; Pred. No. 3.4e-131;
Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1486 TCAGCAGATGCTCAGCCACTGGTAGGAGGAATTTGGTGTACACTTCATCAAAATCT 1545
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96184 TCCACAGATGCTCAGCCACTGGTAGGAGGAATTTGGTGTACACTTCATCAAAATCT 96243

QY 1546 ACCTTTAAACAGAGAGAGAAAGACAGCAGGTATCCAGACCTAGATCTTCAACTCA 1605
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96244 ACCTTTAAACAGAGAGAGAAAGACAGCAGGTATCCAGACCTAGATCTTCAACTCA 96303

QY 1606 CAGGGCAGGAAGTTTATCATGCTTGTGCTGAACCACTCCCAATAGGGGGCCTGAGTAT 1665
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96304 CAGGGCAGGAAGTTTATCATGCTTGTGCTGAACCACTCCCAATAGGGGGCCTGAGTAT 96363

QY 1666 GCAACCCCAATCATGATGATGCTAGGAGCAGCCACAACTTCAGTTGGTCAAGCCCTCC 1725
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96364 GCAACCCCAATCATGATGATGCTAGGAGCAGCCACAACTTCAGTTGGTCAAGCCCTCC 96423

QY 1726 ACATCCACTTTCAGGCTAGGGGAACCAACCTCCCACTAGTGGAACTTCAATACA 1785
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96424 ACATCCACTTTCAGGCTAGGGGAACCAACCTCCCACTAGTGGAACTTCAATACA 96483

QY 1786 CTTCTCTCCAGGACTGACAGTGTCTCTCAGCCAGCCAGCCAGTATGATACCCCGAAGCT 1845
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96484 CTTCTCTCCAGGACTGACAGTGTCTCTCAGCCAGCCAGCCAGTATGATACCCCGAAGCT 96543

QY 1846 GGGAGCCAGGTCTACCTGCCAGACGAATTTGGTGTACAGGTGCCAGAGACACACAA 1905
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96544 GGGAGCCAGGTCTACCTGCCAGACGAATTTGGTGTACAGGTGCCAGAGACACACAA 96603

QY 1906 GAAGTATCAGGAGCAGGAAGGATGGGGAATGTGATCTTTTAAAGAAATCCCTTTGA 1962
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96604 GAAGTATCAGGAGCAGGAAGGATGGGGAATGTGATCTTTTAAAGAAATCCCTTTGA 96660

```

```

RESULT 11
AC025661
LOCUS      Homo sapiens chromosome 3 clone RP11-322A20 map 3, LOW-PASS
DEFINITION      Homo sapiens chromosome 3 clone RP11-322A20 map 3, LOW-PASS
SEQUENCE SAMPLING.
AC025661
VERSION      AC025661.2 GI:7543781
KEYWORDS      HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 146515)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 3, clone RP11-322A20
Unpublished
2 (bases 1 to 146515)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

```

```

O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 146515)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 12, 2000 this sequence version replaced gi:7330271.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7942
Center clone name: 322_A_20
-----
* NOTE: This record contains 164 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 836: contig of 836 bp in length
* 837 936: gap of 100 bp
* 937 1748: contig of 812 bp in length
* 1749 1848: gap of 100 bp
* 1849 2680: contig of 832 bp in length
* 2681 2780: gap of 100 bp
* 2781 3534: contig of 754 bp in length
* 3535 3634: gap of 100 bp
* 3635 4383: contig of 749 bp in length
* 4384 4483: gap of 100 bp
* 4484 5261: contig of 778 bp in length
* 5262 5361: gap of 100 bp
* 5362 6136: contig of 775 bp in length
* 6137 6236: gap of 100 bp

```

6237	7045:	contig of 809 bp	in length
7046	7145:	gap of 100 bp	
7146	7953:	contig of 808 bp	in length
7954	8053:	gap of 100 bp	
8054	8857:	contig of 804 bp	in length
8858	8957:	gap of 100 bp	
8958	9784:	contig of 827 bp	in length
9785	9884:	gap of 100 bp	
9885	10718:	contig of 834 bp	in length
10719	10818:	gap of 100 bp	
10819	11640:	contig of 822 bp	in length
11641	11740:	gap of 100 bp	
11741	12534:	contig of 794 bp	in length
12535	12634:	gap of 100 bp	
12635	13407:	contig of 773 bp	in length
13408	13507:	gap of 100 bp	
13508	14276:	contig of 769 bp	in length
14277	14376:	gap of 100 bp	
14377	15185:	contig of 809 bp	in length
15186	15285:	gap of 100 bp	
15286	16086:	contig of 801 bp	in length
16087	16186:	gap of 100 bp	
16187	16986:	contig of 800 bp	in length
16987	17086:	gap of 100 bp	
17087	17913:	contig of 827 bp	in length
17914	18013:	gap of 100 bp	
18014	18841:	contig of 828 bp	in length
18842	18941:	gap of 100 bp	
18942	19770:	contig of 829 bp	in length
19771	19870:	gap of 100 bp	
19871	20689:	contig of 819 bp	in length
20690	20789:	gap of 100 bp	
20790	21575:	contig of 786 bp	in length
21576	21675:	gap of 100 bp	
21676	22423:	contig of 748 bp	in length
22424	22523:	gap of 100 bp	
22524	23322:	contig of 799 bp	in length
23323	23422:	gap of 100 bp	
23423	24339:	contig of 817 bp	in length
24240	24339:	gap of 100 bp	
24340	25199:	contig of 860 bp	in length
25200	25299:	gap of 100 bp	
25300	26131:	contig of 832 bp	in length
26132	26231:	gap of 100 bp	
26232	27103:	contig of 872 bp	in length
27104	27203:	gap of 100 bp	
27204	28060:	contig of 857 bp	in length
28061	28160:	gap of 100 bp	
28161	29013:	contig of 853 bp	in length
29014	29113:	gap of 100 bp	
29114	29944:	contig of 831 bp	in length
29945	30044:	gap of 100 bp	
30045	30746:	contig of 702 bp	in length
30747	30846:	gap of 100 bp	
30847	31652:	contig of 806 bp	in length
31653	31752:	gap of 100 bp	
31753	32559:	contig of 807 bp	in length
32560	32659:	gap of 100 bp	
32660	33470:	contig of 811 bp	in length
33471	33570:	gap of 100 bp	
33571	34386:	contig of 816 bp	in length
34387	34486:	gap of 100 bp	
34487	35290:	contig of 804 bp	in length
35291	35390:	gap of 100 bp	
35391	36199:	contig of 809 bp	in length
36200	36299:	gap of 100 bp	
36300	37156:	contig of 857 bp	in length
37157	37256:	gap of 100 bp	
37257	38097:	contig of 841 bp	in length
38098	38197:	gap of 100 bp	
38198	39015:	contig of 818 bp	in length
39016	39115:	gap of 100 bp	
39116	39895:	contig of 780 bp	in length

		39896	39995:	gap of	100 bp	
*		39996	40786:	contig of	791 bp	in length
*		40787	40886:	gap of	100 bp	
*		40887	41686:	contig of	800 bp	in length
*		41687	41786:	gap of	100 bp	
*		41787	42543:	contig of	763 bp	in length
*		42550	42649:	gap of	100 bp	
*		42650	43459:	contig of	810 bp	in length
*		43460	43559:	gap of	100 bp	
*		43560	44293:	contig of	734 bp	in length
*		44294	44393:	gap of	100 bp	
*		44394	45225:	contig of	832 bp	in length
*		45226	45325:	gap of	100 bp	
*		45326	46138:	contig of	813 bp	in length
*		46139	46238:	gap of	100 bp	
*		46239	46987:	contig of	749 bp	in length
*		46988	47087:	gap of	100 bp	
*		47088	47853:	contig of	766 bp	in length
*		47854	47953:	gap of	100 bp	
*		47954	48758:	contig of	805 bp	in length
*		48759	48858:	gap of	100 bp	
*		48859	49664:	contig of	806 bp	in length
*		49665	49764:	gap of	100 bp	
*		49765	50565:	contig of	805 bp	in length
*		50570	50669:	gap of	100 bp	

	Query Match	23.6%;	Score 462.8;	DB 2;	Length 146515;
	Best Local Similarity	99.4%;	Pred. No. 6.9e-128;		
	Matches 475;	Conservative	0;	Mismatches 2;	Indels 1; Gaps
QY	1486	TCTGCAGAGTATGCTCAGCCACTGGTAGAGGAATGTTGGTACACTTCACTTCACTCAAGAATCT	1545		
DB	57265	TCCACAGAGTATGCTCAGCCACTGGTAGAGGAATGTTGGTACACTTCACTCAAGAATCT	57324		
QY	1546	ACCTTTAAACCAAGAGAGAAAGAACGAGGCTATGACAGCCTAGATCCCTTACAACCTCA	1605		
DB	57325	ACCTTTAAACCAAGAGAGAAAGAACGAGGCTATGACAGCCTAGATCCCTTACAACCTCA	57384		
QY	1606	CCAGGGCAGGAAGTTTATCATGCCCTATGCTGAACCACTCCCAATTCACGGGGCCTCAGGTAT	1665		
DB	57385	CCAGGGCAGGAAGTTTATCATGCCCTATGCTGAACCACTCCCAATTCACGGGGCCTCAGGTAT	57444		
QY	1666	GCAACCCCAATCATCATGACATGTCAGGGCACCCCACTTCACTAGTTGGTCAGGCCCTCC	1725		
DB	57445	GCAACCCCAATCATCATGACATGTCAGGGCACCCCACTTCACTAGTTGGTCAGGCCCTCC	57504		
QY	1726	ACATCCACTTTCAAGCTTACGGGGAACCAACTCCCCCACTAGTGGGAACCTTACAATACA	1785		
DB	57505	ACATCCACTTTCAAGCTTACGGGGAACCAACTCCCCCACTAGTGGGAACCTTACAATACA	57564		
QY	1786	CTTCTCTCCAGGACTGACAGCTGCTCCTCAGGCCAGGGCCAGTAGTATGATACCCCGAAGCT	1845		
DB	57565	CTTCTCTCCAGGACTGACAGCTGCTCCTCAGGCCAGGGCCAGTAGTATGATACCCCGAAGCT	57624		
QY	1846	GGGAAGCCAGGCTTACCTGCCCCAGACGAATTTGGTGTACCAAGTGCCTACAGAGCACAA	1905		
DB	57625	GGGAAGCCAGGCTTACCTGCCCCAGACGAATTTGGTGTACCAAGTGCCTACAGAGCACAA	57684		
QY	1906	GAAGTATCAGGAGCAGGAAGGATGGGGAATGT-GATGTTTTTAAAGAAATCCTTTGA	1962		
DB	57685	GAAGTATCAGGAGCAGGAAGGATGGGGAATGTGATGTTTTTAAAGAAATCCTTTGA	57742		

RESULT 12

AC126975/c
FOCUS

LOCUS DEFINITION

ACCESSION

ACCESSION
VERSION

KEYWORDS
SOURCE

SOURCE ORGANISM

.....

AC136975 203102 bp DNA linear HTG 11-JUL-2002
Rattus norvegicus clone CH230-234N3, *** SEQUENCING IN PROGRESS
*** 74 unordered pieces.
AC136975
AC136975.1 GI:21731383
HTG; HTGS_PHASE1.
Norway rat.
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 203102)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T., Barabara,J., Benton,J., Bmaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Harris,C., Harris,K., Hart,M., Haviak,P., Hale,S., Hamilton,K., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tanssey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 203102)

Worley,K.C.

Direct Submission

Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GJZQ

Center clone name: CH230-234N3

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 145293 bases at least Q40

Consensus quality: 154399 bases at least Q30

Consensus quality: 161802 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently consists of 74 contigs. The true order of the pieces is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1

1310: contig of 1310 bp in length

1410: gap of unknown length

1411: contig of 1087 bp in length

2498: gap of unknown length

2597: gap of unknown length

2998: contig of 1121 bp in length

3718: gap of unknown length

3819: contig of 1156 bp in length

4974: gap of unknown length

5074: contig of 1044 bp in length

6118: contig of 1058 bp in length

6219: gap of unknown length

7276: gap of unknown length

7376: gap of unknown length

8394: contig of 1018 bp in length

8494: gap of unknown length

8999: contig of 1405 bp in length

9999: gap of unknown length

11010: contig of 1011 bp in length

11110: gap of unknown length

12429: contig of 1319 bp in length

12430: gap of unknown length

12433: contig of 1704 bp in length

14334: gap of unknown length

15599: contig of 1266 bp in length

15699: gap of unknown length

16727: contig of 1027 bp in length

16826: gap of unknown length

18918: contig of 2092 bp in length

19018: gap of unknown length

20837: contig of 1819 bp in length

20937: gap of unknown length

23038: contig of 2099 bp in length

23136: gap of unknown length

24890: contig of 1754 bp in length

24990: gap of unknown length

26683: contig of 1693 bp in length

26783: gap of unknown length

28087: contig of 1304 bp in length

28187: gap of unknown length

29847: contig of 1660 bp in length

31165: gap of unknown length

31265: contig of 1218 bp in length

32327: contig of 1062 bp in length

32427: gap of unknown length

33554: contig of 1127 bp in length

33555: gap of unknown length

34694: contig of 1040 bp in length

34794: gap of unknown length

36507: contig of 1713 bp in length

36608: gap of unknown length

38322: contig of 1715 bp in length

38422: gap of unknown length

40396: contig of 1974 bp in length

40496: gap of unknown length

42695: contig of 2199 bp in length

42696: gap of unknown length

43853: contig of 1058 bp in length

43954: gap of unknown length

45008: contig of 1055 bp in length

45108: gap of unknown length

47026: contig of 1918 bp in length

47126: gap of unknown length

49195: contig of 2069 bp in length

49295: gap of unknown length

51748: contig of 2453 bp in length

51848: gap of unknown length

54202: contig of 2354 bp in length

54302: gap of unknown length

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 472: contig of 472 bp in length
 * 473 572: gap of 100 bp
 * 573 3945: contig of 3373 bp in length
 * 3946 4045: gap of 100 bp
 * 4046 11056: contig of 7011 bp in length
 * 11057 11156: gap of 100 bp
 * 11157 17728: contig of 6572 bp in length
 * 17729 17828: gap of 100 bp
 * 17829 30556: contig of 12728 bp in length
 * 30557 30656: gap of 100 bp
 * 30657 49294: contig of 18638 bp in length
 * 49295 49394: gap of 100 bp
 * 49395 64509: contig of 15115 bp in length
 * 64510 64609: gap of 100 bp
 * 64610 83042: contig of 18433 bp in length
 * 83043 83142: gap of 100 bp
 * 83143 112373: contig of 29231 bp in length
 * 112374 112473: gap of 100 bp
 * 112474 145663: contig of 33190 bp in length
 * 145664 145763: gap of 100 bp
 * 145764 192247: contig of 46484 bp in length.

FEATURES

source
 1. .192247
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="3"
 /chromosome="3"
 /clone_lib="RPC1-11 Human Male BAC"
 /clone="RP11-319J24"
 1. 472
 /note="assembly_fragment
 clone_end:SP6
 vector_side:right"
 573. .3945
 /note="assembly_fragment"
 4046. .11056
 /note="assembly_fragment"
 11157. .17728
 /note="assembly_fragment"
 17829. .30556
 /note="assembly_fragment"
 30657. .49294
 /note="assembly_fragment"
 49395. .64509
 /note="assembly_fragment"
 64610. .83042
 /note="assembly_fragment
 clone_end:T7
 vector_side:left"
 83143. .112373
 /note="assembly_fragment"
 112474. .145663
 /note="assembly_fragment"
 145764. .192247
 /note="assembly_fragment"
 58524 a 38157 c 36840 g 57721 t 1005 others
 ORIGIN

Query Match 11.1%; Score 218; DB 2; Length 192247;
 Best Local Similarity 100.0%; Pred. No. 8.7e-54;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 505 GGATGTTATGGAACACTGGGATGGAGTCGTGGTGTATCGCGGATCCTCAAAATAACAGCA 564
 |||||
 Db 30877 GGATGTTATGGAACACTGGGATGGAGTCGTGGTGTATCGCGGATCCTCAAAATAACAGCA 30818

QY 565 TCATCTCTGCTGGAGTGGAGTACCACACAGGCGACAGAGAACAGTGGAAACCCAAAAA 624
 |||||
 Db 30817 TCATCTCTGCTGGAGTGGAGTACCACACAGGCGACAGAGAACAGTGGAAACCCAAAAA 30758

QY 625 GCCAGGCTGAAAAAACCTGGACCGCTTGGCTGCTTTGCCACTGATGAATACCACTGG 684
 |||||
 Db 30757 GCCAGGCTGAAAAAACCTGGACCGCTTGGCTGCTTTGCCACTGATGAATACCACTGG 30698
 |||||
 QY 685 TTACAAATAGATTGAATAGGAAAAAATAACAGG 722
 |||||
 Db 30697 TTACAAATAGATTGAATAGGAAAAAATAACAGG 30660
 |||||
 RESULT 14
 AXI18816
 LOCUS AXI18816
 DEFINITION Sequence 1 from Patent WO0129219.
 ACCESSION AXI18816
 VERSION AXI18816.1 GI:14035774
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1464)
 AUTHORS Donoho,G., Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and
 Sands,A.T.
 TITLE Human cub-domain-containing protein and gene encoding the same
 JOURNAL Patent: WO 0129219-A 1 26-APR-2001;
 Lexicon Genetics Incorporated (US)
 FEATURES
 source 1. .1464
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 440 a 303 c 359 g 362 t
 ORIGIN
 Query Match 10.4%; Score 204; DB 6; Length 1464;
 Best Local Similarity 54.1%; Pred. No. 7.3e-50;
 Matches 476; Conservative 0; Mismatches 380; Indels 24; Gaps 2;
 QY 121 AATGAATCACATTGCTGTTTCATGAGTGGAAATCCATGTTTCTGACGCGGATTTTGGCC 180
 |||||
 Db 220 AGTGAAGTAACCGTCGCTTTGAGATGGATCCCAATTTCTGGCGGGGTTTTTGTCTG 279
 |||||
 QY 181 TCATACCTCTGTTATAGATAAACAAGATCTAATTACTTGTGTTGGACACTGCATCAATTTT 240
 |||||
 Db 280 ACCTATGCCGAGCAGCGACCATCCAGATTTAATAACATGTTTGGACAGAGTACCATAT 339
 |||||
 QY 241 TTGGAACCTGAGTTTCAGTAAGTACTGCCAGCTGGTTGCTGCTTCCTTTGTTGTAGATA 300
 |||||
 Db 340 TTGAAGACAGAAATACAGCAAAATTCGCCAGCTGGTTGTAGAGACGTAGCAGGAGACATT 399
 |||||
 QY 301 TCTGGAAACAATTCCTCATGGATATAGAGATTCCTGCCCATTTGCGATGGCTGGTGTGCAT 360
 |||||
 Db 400 TCTGGGAATATGTTAGATGATATAGAGATACCTCTTTATTTGCAAGGCTGCATCCT 459
 |||||
 QY 361 GCAGGAGTAGTGTCAAACACAGCTTGGCGGCAATCAGTGTGTTAATTAGTAAAGGTATT 420
 |||||
 Db 460 GCAGGAATAATTGCTGATGNACTAGTGGCCAGATCAGTGTGTTTACGCCAAGGGATC 519
 |||||
 QY 421 CCTATTATGAAAGTCTTTTGGCTAAACACGTCCACATCTGTTGGTGAGCACTTATCTACA 480
 |||||
 Db 520 AGTCGATATGAAGGGATTCTGGCCAATGGTGTCTTTCGAGGGATGTTCCCTGTCCAGAC 579
 |||||
 QY 481 AGTCTTTTACATTTTAAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGGTGTG 540
 |||||
 Db 580 AAGCGATTCTGTTTACCTCCAAATGGTTGCAGCAGATCCCTTTGAGTTTGAACCTGACGGG 639
 |||||
 QY 541 ATCGCGGATCCTCAATAACAGCATCATCTGCTGGAGTGACTGACCAACACAGGSCAA 600
 |||||
 Db 640 -----CAAAATCAGAGCTTCTTCTCATGGCAGTCGGTCAATGAGATGGAGAC 687
 |||||
 QY 601 GAGAACAGTTGGAACCCCAACAGCGCTGAAACAACTGGACCGCTTGGGCTGCT 660
 |||||
 Db 688 CRAGTTCACTGGTCTCTCTGCGCCAAAGCCGACTTCAGGACCAAGGCCCATCATGSGCTTCG 747
 |||||

```
QY 661 TTTGCCACTGATGAATACC-----AGTGTTTACAATAGATTGTAATAGGAA 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 748 GCGCAGTAGCAACAACCAACACAGAGAGTGGCTGGAGATCGATTGGGGGAGAA 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 709 AAGAAAATACAGGAGCATATATACCACTGGATCCACCATGGTGGAGCACAATTACTATGTG 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 808 AAGAAAATACAGGAAATTAGGACCCACAGGATCTACACAGTGAACCTTCAACTTTTATGTT 867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 769 TCTGCTACAGAAATCTGTACAGTGTATGGCGAGAAATGACTGTGTACAGAGAGCCT 828
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 868 AAGAGTTTGTGATGACTTCAAAACAATAATTTCTAAGTGGAGACCTATAAAGGAATT 927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 829 GGTGTGGAGCAAGATTAAGATATTTCAAGGAAACAAAGATTTATCACAGGATGTGCGTAAT 888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 928 GTGAATAATGAAGAAAGGTGTTTCAAGGTAACTCTAATCTTTCGGGACCCAGTGCAAAAC 987
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 889 AACTTTTCCCAACCAATTAATGACGCTTTTATAGAGTGAATCCCTACCCAATGGCAGCAG 948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 988 AATTTTCATCCCTCCCATCTGTGGCCAGATATGTGCGGGTGTGCCCCAGACATGGCACCAG 1047
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 949 AAAATTGCCATGAATGAGCTGCTCGGATGTCAGTTTA 988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1048 AGGATAGCCTTGAAGGTGGAGCTCATTTGGTGGCAGATTA 1087
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AX118820 AX118820 1620 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 5 from Patent WO0129219.
ACCESSION AX118820
VERSION AX118820.1 GI:14035776
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1620)
AUTHORS Donoho,G., Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and
Sands,A.T.
TITLE Human cub-domain-containing protein and gene encoding the same
JOURNAL Patent: WO 0129219-A 5 26-APR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
    source
        1..1620
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 456 a 353 c 422 g 389 t
ORIGIN
    Query Match 10.4%; Score 204; DB 6; Length 1620;
    Best Local Similarity 54.1%; Pred. No. 7.5e-50;
    Matches 476; Conservative 0; Mismatches 380; Indels 24; Gaps 2;

QY 121 AATGAATACATTTGCTGTTTCATGAGTGGAAATCCATGTTTCTGGACCGGATTTTGGCC 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 AGTGAAGTACCGCTCGCTTTGAGAGTGGATCCACATTTCTGGCCGGGTTTTTGGCTG 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 TCATCTCTGTTATAGTAACAAGATCTAATTACTTGTGTTGGACACTGTCACCAATTTT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 ACCTATGCGAGCAGCGACCATCCAGATTTTAATACATGTTTGGAAACGAGTAGCCATTAT 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 TTGGAACCTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTGCTTCCTTTTGTGGAGATA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 TTGAGACACAATACAGCAATTTCTGCCAGCTGGTGTAGAGACGTAGCAGGAGACATT 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 TCTGGAAACAATCTCTATGAGATATAGAGATTCCTCGCCATTTGTCATGGCTGTGTCAT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 TCTGGGAATATGGTAGATAGATATAGATACCTCTTTATTTGTGCAAGCTGCCATCCAT 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 GCAGGAGTAGTCAACACAGCTTGGGGCGCCCAATCATAGTGTGTAATAGTAAGGTATT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 GCAGGAATAATTGCTGATGAACATAGTGGCCAGATCATGCTGTGCTTACGCCGCAAGGATC 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: January 21, 2003, 07:26:07
Job time : 5794.22 secs

```
QY 421 CCTATTATTAAGATTTCTTTTGGCTAACAAACGTCACATCTGTGGTGGGACACACTATCTACA 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 676 AGTCGATATGAAGGGATTTCTGGCCATGTTCTTTTCGAGGATGGTTTCCCTGTCCAGAC 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 481 AGTCTTTTACATTTTAAGACAAGTGTATTTAGAACACTGGGATGGGATGGAGTCTGTTG 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 736 AAGCGATTTCTGTTTACCTCCCAATGTTGCAGCAGATCTTGTAGTTTGAACCTGACGGG 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 541 ATCGCGGATCCTCAAAATACAGCATCATCTGTCTGGAGTGGACTGACCAACAGAGGCA 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 796 -----CAAATCAGAGCTTCTTCTCATGCGAGTCTGCTCAATGAGAGTGGAGAC 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 601 GAGAACAGTTTGAACCCCAAAAGCCAGCTGAAAAACCTTGGAGCCGCTTGGGCTGCT 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 844 CRAAGTTCACTGGTCTCTCTGGCCAAAGCCGACTTCAGGACCAAGGCCCATCATGGGCTCG 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 661 TTTGCCACTGATGAATACC-----AGTGTTTACAATAGATTGTAATAGGAA 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 904 GCGCAGTAGCAACAACCAACCAACAGAGAGTGGCTGGAGATCGATTTGGGGGAGAA 963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 709 AAGAAAATACAGGCAATTTATACCACTGGATCCACCATGGTGGAGCACAATTTACTATGTG 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 964 AAGAAAATACAGGAAATTAGGACCCACAGGATCTACACAGTCCGAACCTTCAACTTTTATGTT 1023
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 769 TCTGCCCTACAGAATCCTGTACAGTGTATGGCGAGAAATGGAGTGTGTACAGAGAGCCT 828
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1024 AAGAGTTTGTGATGAACITCAAAAACAATAATTTCTAAGTGAAGACCTATAAAGGAATT 1083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 829 GGTGTGGAGCAAGATAGATATTTCAAGGAAACAAAGATTTATCACAGGATGTGCGTAAT 888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 GTGAATAATGAAGAAAGGTGTTTTCAGGGTAACTCTAACTTTCCGGACCCAGTGCAGAAAC 1143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 889 AACTTTTGGCCACCAATTTATGCAACGTTTTATTAGAGTGAATCCTTACCCCAATGGCAGCAG 948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1144 AATTTCATCCTCCCATCTGTCGGCCAGATATGTCGGGGTGTGTCGCCAGACATGGCACCAG 1203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 949 AAAATTGCCATGAATGAGCTGCTCGGATGTCAGTTTA 988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1204 AGGATAGCCTTGAAGGTGGAGCTCATTTGGTGGCAGATTA 1243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 01:52:19 ; Search time 75.5882 Seconds
(without alignments)
9217.500 Million cell updates/sec

Title: US-10-060-830-1
Perfect score: 2280
Sequence: 1 gccgcgccccgcctgggc.....tttgaagatgatgctgttt 2280

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	544.6	23.9	640	4	US-09-328-111-480
2	91.6	4.0	6585	3	US-08-746-111-4
3	86.6	3.8	6909	2	US-08-804-196-1
4	86.6	3.8	6909	2	US-08-658-340-1
5	86.6	3.8	6909	3	US-08-746-111-26
6	86.4	3.8	2303	2	US-08-480-229C-9
7	86.4	3.8	2303	2	US-08-659-235C-9
8	78.6	3.4	318	2	US-08-480-229C-19
9	78.6	3.4	318	2	US-08-659-235C-19
10	78.2	3.4	1780	2	US-08-480-229C-11
11	78.2	3.4	1780	2	US-08-659-235C-11
12	76.8	3.4	1384	1	US-07-607-538C-1
13	76.8	3.4	1384	2	US-08-162-402B-1
14	76.8	3.4	1934	2	US-08-162-402B-7
15	63.4	2.8	4629	2	US-08-484-891-7
16	63.4	2.8	4670	3	US-08-717-294-41
17	63.4	2.8	4999	4	US-09-470-618-14
18	63.4	2.8	4999	4	US-09-364-862-14
19	63.4	2.8	5035	2	US-08-882-083-1
20	63.4	2.8	5035	2	US-08-558-107-1
21	63.4	2.8	5035	3	US-09-243-539-1
22	63.4	2.8	6999	1	US-08-276-594A-1
23	63.4	2.8	7056	1	US-08-121-202-1
24	63.4	2.8	7493	1	US-08-212-133A-7
25	63.4	2.8	7493	1	US-08-474-503-5
26	63.4	2.8	7493	2	US-08-670-707A-5
27	63.4	2.8	7493	4	US-09-037-601-5

28	63.4	2.8	7493	4	US-09-315-179-5	Sequence 5, Appl
29	63.4	2.8	7493	5	PCT-US94-13200-5	Sequence 5, Appl
30	63.4	2.8	8241	6	5171844-1	Patent No. 5171844
31	63.4	2.8	8967	1	US-08-366-851A-1	Sequence 1, Appl
32	63.4	2.8	9009	1	US-07-864-004B-3	Sequence 3, Appl
33	63.4	2.8	9009	1	US-08-251-937A-3	Sequence 3, Appl
34	63.4	2.8	9009	1	US-08-212-133A-1	Sequence 1, Appl
35	63.4	2.8	9009	1	US-08-474-503-1	Sequence 1, Appl
36	63.4	2.8	9009	2	US-08-670-707A-1	Sequence 1, Appl
37	63.4	2.8	9009	4	US-09-037-601-1	Sequence 1, Appl
38	63.4	2.8	9009	4	US-09-315-179-1	Sequence 1, Appl
39	63.4	2.8	9009	4	US-09-523-656-1	Sequence 1, Appl
40	63.4	2.8	9009	5	PCT-US93-03275-3	Sequence 3, Appl
41	63.4	2.8	9009	5	PCT-US94-13200-1	Sequence 1, Appl
42	63.4	2.8	9354	1	US-08-683-839B-2	Sequence 2, Appl
43	63.4	2.8	11933	4	US-09-470-618-13	Sequence 13, Appl
44	63.4	2.8	11933	4	US-09-364-862-13	Sequence 13, Appl
45	63.2	2.8	7032	4	US-09-324-867-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-328-111-480/c
; Sequence 480, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Casino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-480

Query Match	23.9%	Score 544.6;	DB 4;	Length 640;
Best Local Similarity	96.7%	Pred. No. 8.5e-134;		
Matches 616;	Conservative 0;	Mismatches 15;	Indels 6;	Gaps 6;
Qy 1196	TTTTGCCCAACCAATATTTCACGCTTTTATTAGAGTGAATCTTACCAATGCCAGCAGAAA	1255		
Db 632	TTTTGCCCAACCAATATTTCACGCTTTTATTAGAGTGAATCTTACCAATGCCAGCAG-AAA	576		
Qy 1256	TTGCCATGAAATGAGCTGCTCGGATGTCAGTTATTCCTAAAGGTGCTCTCCAAAC	1315		
Db 575	TTGCCATGAAATGAGCTGCTCGGATGTCAGTTATTCCTAAAGGTGCTCTCCAAAC	518		
Qy 1316	TTACTTAACCTCCACCTCCTCGGAACACAGCAATGACCTCAAAAACACTACGCCCTCCAA	1375		

```

; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6...6554
US-08-746-111-4

Query Match      4.0%; Score 91.6; DB 3; Length 6585;
Best Local Similarity 53.3%; Pred. No. 6.6e-17;
Matches 193; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

Qy 953 CTTGGGTGTCCTTTGGCCACTGATGAATACCAAGTGGTWTACAAATAGATTGTAATAGGAAA 1012
    |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6214 CTTGGCAAGCCAGCAACCAACCAACACGACGAGTGGTTACAAAGTCGTGCTGCTAAAAATCA 6273

Qy 1013 AGAAATAACAGGATATTATACCACCTGGATCCACCATGGTGGAGCACAAATTACTATGTGT 1072
    |||| |||| ||| ||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6274 AGAAGGTAAACGGCCATCGTAAACGAGGCGCTGTAACTCTCTCTCTGAGATGTACGTGA 6333

Qy 1073 CTGGCTACAGAAATCCTGTACACAGTGATGGCGACAAAATGGACTGTGTACAGAGAGCCTG 1132
    ||||| |||| ||| |||| |||| ||| ||| |||| ||| ||| |||| ||| ||| |||
Db 6334 AGAGCTACAGCATCCAGTACAGTACCAGGGTGTGGCATGGAACCTTTACCGACAGAAAT 6393

Qy 1133 GTGTGGACAAAGATAAGATATTTCAGGAAACAAAGATTATCACAGAGATGTGGCTAATA 1192
    ||| |||| |||| |||| |||| |||| ||| ||| |||| ||| ||| |||| ||| |||
Db 6394 CCTCATGGTGGCAAGATATTTTGAAGGAAACAGCAATACCAAGGGGCACATGAAGAACT 6453

Qy 1193 ACTTTTGGCCACCATTATTCGACGTTTTTATTAGAGTGAATCCTTACCCAATGGCAGCAGA 1252
    || || |||| |||| ||| |||| ||| |||| ||| |||| ||| |||| ||| ||||
Db 6454 TTTTCAACCCGCCCATTTATTTCCAGATTATTTCCGCATCATTTCTTAAAAACATGGGAACCCAGA 6513

Qy 1253 AAATTGCCATGAAAATGGAGCTGCTCGGATGTGCAGTTTTATTCTTAAAGGTGCTCTCTCCAA 1312
    || |||| ||| ||| |||| |||| ||| |||| ||| |||| ||| |||| ||| |||
Db 6514 GCATCGCCCTTCGCCTAGAGCTCTTCGGCTGTGACATTTATTAGAAATTAAATTCCTCAAAA 6573

Qy 1313 AA 1314
    ||
Db 6574 AA 6575

```

```

1 RESULT 3
2 US-08-804-196-1
3 : Sequence 1, Application US/08804196
4 : Patent No. 5874256
5 : GENERAL INFORMATION:
6 : APPLICANT: Bertina, Rogier
7 : APPLICANT: Reitsma, Pieter
8 : TITLE OF INVENTION: A method for diagnosing an increased
9 : TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
10 : TITLE OF INVENTION: thrombosis and kit for use with the same.
11 : NUMBER OF SEQUENCES: 14
12 : CORRESPONDENCE ADDRESS:
13 : ADDRESSEE: Akzo No. 5874256el Patent Department
14 : STREET: 1300 Piccard Drive, Suite 206
15 : CITY: Rockville
16 : STATE: Maryland
17 : COUNTRY: USA
18 : ZIP: 20850
19 : COMPUTER READABLE FORM:
20 : MEDIUM TYPE: Floppy disk
21 : COMPUTER: IBM PC compatible
22 : OPERATING SYSTEM: PC-DOS/MS-DOS
23 : SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
24 : CURRENT APPLICATION DATA:
25 : APPLICATION NUMBER: US/08/804,196
26 : FILING DATE:
27 : PRIOR APPLICATION DATA:
28 : APPLICATION NUMBER: US 08/454,353
29 : FILING DATE: 06-JUN-1995
30 : ATTORNEY/AGENT INFORMATION:
31 : NAME: Gormley, Mary E.
32 : REGISTRATION NUMBER: 34,409

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 948-7400
TELEFAX: (301) 948-9751
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6909 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-804-196-1

Query Match 3.8%; Score 86.6; DB 2; Length 6909;
Best Local Similarity 51.7%; Pred. No. 2.2e-15;
Matches 259; Conservative 0; Mismatches 224; Indels 18; Gaps 2;

QY 796 TTAAAGACAGTGGATGTTATGGAACACATGGGATGGAGTCTGCTGTGATCGGGATCCT 855
DB 6271 TGTAGGTAATGGATGTTCCACACCCCTGGGTATGGAATGGAAGATAGAAAACAAG 6330
QY 856 CAAATAACACATCATCTCTGTGGAGTGGACTGACACACAGGGCAAGAGAACAGTTGG 915
DB 6331 CAAATCACAGCTCTTC-----GTTTAAAGAAATCTTGGTGGGAGATTACTGG 6378
QY 916 AARCCCAAAAGCCAGGCTGAAAAACCTGGAC-----CGCCTTGGGCTGCTTTGGCC 969
DB 6379 GAACCCCTTCGTCGCCCTCTGAATGTCAGGAGCTGTGAATCGCTGGCAAGCAAGGCA 6438
QY 970 ACTGATGAATACACAGTGGTTACAAATAGATTGTAATAAGGAAAGAAATAACAGGCATT 1029
DB 6439 AACAACAATAAGCAGTGGCTAGAAATGTACTCAAGATCAAGAGATAACCGCAATT 6498
QY 1030 ATAACACATGGATCCACCATGGTGGAGCACAAATTACTATGTCTGTCTACAGAACTCTG 1089
DB 6499 ATAACACAGGCTGCAAGTCTCTGCTCTGAAATGTATGTAAGAGCTATACCATCCAC 6558
QY 1090 TACAGTGTATGGCAGAAATGGACTGTGTACAGAGAGCCTGGTGGGAGCAAGATAAG 1149
DB 6559 TACAGTGAGCAGGAGTGGAAATGGAACCATACAGGCTGAAATCCTCATGTTGGGACAAG 6618
QY 1150 ATATTTCAAGGAACAAGATTATACAGGATGTCGCTAATTAACCTTTTGGCCACCAATT 1209
DB 6619 ATTTTGAAGGAATACTAATACCAAGGACATGTGAAGAACTTTTCAACCCCCCAATC 6678
QY 1210 ATTCGACGTTTATTAGATGAATCCTACCAATGGCAGAGAAATTTGCCATGAAAATG 1269
DB 6679 ATTTCCAGGTTTATCCGCTGTCATCTCTAACAACATGGAATCAAGTATTACACITCGGCTG 6738
QY 1270 GAGCTGCTGGATGTCAGTTT 1290
DB 6739 GAACTCTTTGGCTGTGATATT 6759

RESULT 4
US-08-658-340-1
Sequence 1, Application US/08658340
Patent No. 5910576
Patent No. 5910576 5861489
GENERAL INFORMATION:
APPLICANT: Bertina, Rogier
APPLICANT: Reitsma, Pieter
TITLE OF INVENTION: A method for diagnosing an increased
TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
TITLE OF INVENTION: thrombosis and kit for use with the same.
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Akzo No. 5910576 5861489el Patent Department
STREET: 1300 Piccard Drive, Suite 206
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,340
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/454,353
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 948-7400
TELEFAX: (301) 948-9751
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6909 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-658-340-1

Query Match 3.8%; Score 86.6; DB 2; Length 6909;
Best Local Similarity 51.7%; Pred. No. 2.2e-15;
Matches 259; Conservative 0; Mismatches 224; Indels 18; Gaps 2;

QY 796 TTAAAGACAGTGGATGTTATGGAACACATGGGATGGAGTCTGCTGTGATCGGGATCCT 855
DB 6271 TGTAGGTAATGGATGTTCCACACCCCTGGGTATGGAATGGAAGATAGAAAACAAG 6330
QY 856 CAAATAACACATCATCTCTGTGGAGTGGACTGACACACAGGGCAAGAGAACAGTTGG 915
DB 6331 CAAATCACAGCTCTTC-----GTTTAAAGAAATCTTGGTGGGAGATTACTGG 6378
QY 916 AARCCCAAAAGCCAGGCTGAAAAACCTGGAC-----CGCCTTGGGCTGCTTTGGCC 969
DB 6379 GAACCCCTTCGTCGCCCTCTGAATGTCAGGAGCTGTGAATCGCTGGCAAGCAAGGCA 6438
QY 970 ACTGATGAATACACAGTGGTTACAAATAGATTGTAATAAGGAAAGAAATAACAGGCATT 1029
DB 6439 AACAACAATAAGCAGTGGCTAGAAATGTACTCAAGATCAAGAGATAACCGCAATT 6498
QY 1030 ATAACACATGGATCCACCATGGTGGAGCACAAATTACTATGTCTGTCTACAGAACTCTG 1089
DB 6499 ATAACACAGGCTGCAAGTCTCTGCTCTGAAATGTATGTAAGAGCTATACCATCCAC 6558
QY 1090 TACAGTGTATGGCAGAAATGGACTGTGTACAGAGAGCCTGGTGGGAGCAAGATAAG 1149
DB 6559 TACAGTGAGCAGGAGTGGAAATGGAACCATACAGGCTGAAATCCTCATGTTGGGACAAG 6618
QY 1150 ATATTTCAAGGAACAAGATTATACAGGATGTCGCTAATTAACCTTTTGGCCACCAATT 1209
DB 6619 ATTTTGAAGGAATACTAATACCAAGGACATGTGAAGAACTTTTCAACCCCCCAATC 6678
QY 1210 ATTCGACGTTTATTAGATGAATCCTACCAATGGCAGAGAAATTTGCCATGAAAATG 1269
DB 6679 ATTTCCAGGTTTATCCGCTGTCATCTCTAACAACATGGAATCAAGTATTACACITCGGCTG 6738
QY 1270 GAGCTGCTGGATGTCAGTTT 1290
DB 6739 GAACTCTTTGGCTGTGATATT 6759

RESULT 5
US-08-746-111-26
Sequence 26, Application US/08746111
Patent No. 6066778
GENERAL INFORMATION:
APPLICANT: Ginsburg, David
APPLICANT: Cui, Jisong

Db 1326 GATTGGAGCCAGAGTACATAAAATCCATACAAAATTCCTACAGCAATGACGGGAAGAC 1385
 QY 1110 ATGAGTGTGTACAGAGAGCCTGTGTGGAGCAAGATAAGATATTTCAAGGAACAAGA 1169
 Db 1386 CTGGCAATGTACAAAGTAAGAGCCAAATGAAGAGATGCTCTTCGTGGAATGTTGA 1445
 QY 1170 TTATCACCAGAGTGTGGTAAATACATTTTGGCCACCAATTTATGCCAGTTTATTAGAT 1229
 Db 1446 TAACAACACACCATATGCTAAATCTTTACACCCCAATCAAGCTCAGTATGTAAGACT 1505
 QY 1230 GAATCTGTACCAATGGCAGCAGAAAATGGCAATGAGAGTGTGGATGTGAGT 1289
 Db 1506 CTACCCCAATTTGTGCAAGGCAATGTAAGAAATGGAATCTTCTGGCTGTGAGCT 1565

RESULT 7
 US-08-659-235C-9
 ; Sequence 9, Application US/08659235C
 ; Patent No. 5877281
 ; GENERAL INFORMATION:
 ; APPLICANT: Quertermous, Thomas
 ; APPLICANT: Hogan, Brigid
 ; APPLICANT: Snodgrass, H. Ralph
 ; APPLICANT: Zupancic, Thomas J.
 ; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
 ; TITLE OF INVENTION: CELL LOCUS-1
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds LLP
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/659,235C
 ; FILING DATE: 05-JUN-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Poissant, Brian M.
 ; REGISTRATION NUMBER: 28,462
 ; REFERENCE/DOCKET NUMBER: 8907-0034-999
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864/9741
 ; TELEX: 66141 Pennie
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2303 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: CDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 619..2058
 ; US-08-659-235C-9

Query Match 3.8%; Score 86.4; DB 2; Length 2303;
 Best Local Similarity 50.2%; Pred. NO. 1.2e-15;
 Matches 241; Conservative 0; Mismatches 236; Indels 3; Gaps 1;
 QY 810 ATGTTATGAACACTGGGAGTGGTGTGATCGGGATCCCTCAATAACAGCATC 869
 Db 1089 ATGCTGTGGGCACTTGGGAATGAAGGTGGATCATATTAATCAAGCAATCACAGCTTC 1148
 QY 870 ATCTGTGTGGAGTGGACTGACCACACAGGCAAGAGAACTGTGGAACCCCAAAAAGC 929

Db 1149 ATCTAATC---ACCGAGCTCTTTTGGACTCCAGAGTGGTATCCCTACTATCTCGACT 1205
 QY 930 CAGGCTGAAAAAACCTGGACCGCTTGGGCTGTCTTTTCCCACTGATGAATACACAGTGGTT 989
 Db 1206 TAATAAGAGGGGCTTATAAATGCTTGGACAGCTGTGTAATAATGACAGATGGCCATGGAT 1265
 QY 990 ACAAAATAGATTTGAATAAGGAAAAAATAACAGGCAATTAACCACTGGATCCACCAT 1049
 Db 1266 TCAGATAAATTTGCAAGAAAAATGAGAGTCATCTGGTGTATTATCCCAAGAGCAAAAAG 1325
 QY 1050 GGTGGAGCACAATTTACTATGTCTGCTACAGAAATCCTTACAGTGTATGATGGGCAGAA 1109
 Db 1326 GATTGGAAGCCAGAGTACATAAATCTTCAAAAATTTGCTACAGCAATGACGGGAAGAC 1385
 QY 1110 ATGAGTGTGTACAGAGCGCTGTGTGGAGCAAGATAAGATATTTCAAGGAACAAGA 1169
 Db 1386 CTGGCAATGTACAAAGTAAGAGCCAAATGAAGAGATGCTCTTCTGGTGAATGTTGA 1445
 QY 1170 TTATCACCAGGATGTGCTAATAACTTTTGGCCACCAATTTATGCAGCTTTTATTAGAT 1229
 Db 1446 TAACAACACACCATATGCTAATCTTTTCACACCCCAATCAAGCTCAGTATGTAAGACT 1505
 QY 1230 GAATCTGTACCAATGGCAGCAGAAAATGGCAATGAGAGTGTGGATGTGAGT 1289
 Db 1506 CTACCCCAATTTGTGCAAGGCAATGTAAGAAATGGAATCTTCTGGCTGTGAGCT 1565

RESULT 8
 US-08-480-229C-19
 ; Sequence 19, Application US/08480229C
 ; Patent No. 5874562
 ; GENERAL INFORMATION:
 ; APPLICANT: Quertermous, Thomas
 ; APPLICANT: Hogan, Brigid
 ; APPLICANT: Snodgrass, H. Ralph
 ; APPLICANT: Zupancic, Thomas J.
 ; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
 ; TITLE OF INVENTION: CELL LOCUS-1
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds LLP
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/480,229C
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Poissant, Brian M.
 ; REGISTRATION NUMBER: 28,462
 ; REFERENCE/DOCKET NUMBER: 8907-0026-999
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864/9741
 ; TELEX: 66141 Pennie
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 318 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-480-229C-19

Query Match 3.4%; Score 78.6; DB 2; Length 318;

[illegible]

```

RESULT 9
US-08-659-235C-19
; Sequence 19, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quettermous, Thomas
; APPLICANT: Hogan, Bridg
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,235C
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0034-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-659-235C-19

```

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```
; NAME/KEY: CDS
; LOCATION: 1..1779
US-08-480-229C-11

Query Match      3.4%; Score 78.2; DB 2; Length 1780;
Best Local Similarity 53.4%; Pred. No. 3.1e-13;
Matches 164; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 980 ACCAGTGGTTACAAATAGATTGTAATGAAGGAAAGAAATACAGGCAATATACCACTG 1039
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 824 AGCGGTGGATTACAGATAAATTTGCAAGAAAGAAATGAGAGTTACTGGTGTGATTACCCAAG 883

QY 1040 GATCCACCATGGTGGAGCACAATTAATGTCTGCTGCTACAGAAATCCTGTACAGTGTG 1099
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 884 GGGCCAGAGGATTTGGAAGCCAGAGTATATAAAATTTCTACAAATTTGCCACAGTAATG 943

QY 1100 ATGGGAGAGAAATGACTGTGTACAGAGAGCTGTGTGGAGCAAGATAAGATATTTCAAG 1159
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 944 ATGGAAGACTTGGGCAATGTACAAAGTGAAGGCCCAATGAAGACATGGTGTTCGTG 1003

QY 1160 GAAACAAAGATTTATCACCAGGATGTGCGTAATAACTTTTGGCCACCAATATTTCACGTT 1219
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1004 GAAACATTGATACACACTCCATATGCTAACTTTTCACACCCCCCATAAAGCTCAGT 1063

QY 1220 TTATTAGAGTAATCCTACCAATGGCAGAGAAATTTGCCATGAAATGGAGCTGCTCG 1279
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1064 ATGAAGACTCTATCCCAAGTTTGTGGAAGACATTCGACTTTTGGGAATGGAATCTTCTG 1123

QY 1280 GATGTCA 1286
   |||||
Db 1124 GCTGTGA 1130

RESULT 11
US-08-659-235C-11
; Sequence 11, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quertemous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pennle & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,235C
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0034-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1779
US-08-659-235C-11

Query Match      3.4%; Score 78.2; DB 2; Length 1780;
Best Local Similarity 53.4%; Pred. No. 3.1e-13;
Matches 164; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 980 ACCAGTGGTTACAAATAGATTGTAATGAAGGAAAGAAATACAGGCAATATACCACTG 1039
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 824 AGCGGTGGATTACAGATAAATTTGCAAGAAAGAAATGAGAGTTACTGGTGTGATTACCCAAG 883

QY 1040 GATCCACCATGGTGGAGCACAATTAATGTCTGCTGCTACAGAAATCCTGTACAGTGTG 1099
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 884 GGGCCAGAGGATTTGGAAGCCAGAGTATATAAAATTTCTACAAATTTGCCACAGTAATG 943

QY 1100 ATGGGAGAGAAATGACTGTGTACAGAGAGCTGTGTGGAGCAAGATAAGATATTTCAAG 1159
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 944 ATGGAAGACTTGGGCAATGTACAAAGTGAAGGCCCAATGAAGACATGGTGTTCGTG 1003

QY 1160 GAAACAAAGATTTATCACCAGGATGTGCGTAATAACTTTTGGCCACCAATATTTCACGTT 1219
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1004 GAAACATTGATACACACTCCATATGCTAACTTTTCACACCCCCCATAAAGCTCAGT 1063

QY 1220 TTATTAGAGTAATCCTACCAATGGCAGAGAAATTTGCCATGAAATGGAGCTGCTCG 1279
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1064 ATGAAGACTCTATCCCAAGTTTGTGGAAGACATTCGACTTTTGGGAATGGAATCTTCTG 1123

QY 1280 GATGTCA 1286
   |||||
Db 1124 GCTGTGA 1130

RESULT 12
US-07-607-538C-1
; Sequence 1, Application US/07607538C
; Patent No. 5455031
; GENERAL INFORMATION:
; APPLICANT: Ceriani Dr., Roberto L.
; APPLICANT: Peterson Dr., Jerry A.
; APPLICANT: Larocca, David J.
; TITLE OF INVENTION: POLYPEPTIDE WITH 46
; TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
; TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES, POLYRIBO-
; TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-
; TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
; TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
; TITLE OF INVENTION: USE THEREOF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: V. Anzel & Assoc.
; STREET: 2055 No. 5455031th Broadway
; CITY: Walnut Creek
; STATE: California
; COUNTRY: USA
; ZIP: 94596
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/607,538C
; FILING DATE: 01-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Viviana Anzel
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: CRFC-004
```



```

APPLICATION NUMBER: US/08/484,891
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/218,335
FILING DATE: 25-MAR-1994
APPLICATION NUMBER: 08/074,920
FILING DATE: 10-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Olstein, Elliot M.
REGISTRATION NUMBER: 24,025
REFERENCE/DOCKET NUMBER: 271010-273
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4629 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA primer
FEATURE:
NAME/KEY: Factor VIII cDNA with
NAME/KEY: B domain deleted
US-08-484-891-7

Query Match      2.8%; Score 63.4; DB 2; Length 4629;
Best Local Similarity 51.6%; Pred. No. 1.6e-08;
Matches 173; Conservative 0; Mismatches 156; Indels 6; Gaps 1;

QY 953 CTGGGCTGTCTTTGCCACTGATGAATACCATCGTGGTTACAATAAGTTGAATAGGAAA 1012
DB 4070 CTTGGAGACCTCAGGTGAATAATCCAAAGAGTGCGTCGCAAGTGGACTTCCAGAAGCAA 4129
QY 1013 AGAATAACAGGCATTAATACCACCTGGATCCACCATGGTGAGCACAAATTAATGTGT 1072
DB 4130 TGAAGTCACAGGAGTAATCTCTCAGGAGTAATAATCTCTGTTACCAGCATGTATGTA 4189
QY 1073 CTGCTACAGAACCTCTACAGTATGGGTCAGAAATGGACTGTGTACAGAGAGCCTG 1132
DB 4190 AGGAGTTCCTCATCTCCAGCAGTCAAGATGGCCATCAGTGGACTCTCTTTTTTCAGAA 4249
QY 1133 GTGTGGAGCAAGATAAGATATTTCAAGBBAACAAGATTATCACCAGGATGTGCGTAATA 1192
DB 4250 GCA-----AAGTAAAGGTTTTTTCAGGGAATCAAGACTCCCTCACACCTGTGGTGA 4303
QY 1193 ACTTTTGGCCACCAATTTATTGCACGTTTTATTAGAGTGAATCCTACCCCAATGGCAGCA 1252
DB 4304 CTCTAGACCCACCGTTACTGCTGCTGCTTACCTTCGATTCCACCCCAGAGTTGGTGCAC 4363
QY 1253 AAATTGCCATGAAAAATGGAGCTGCTCGGATGTCAG 1287
DB 4364 AGATTGCCCTGAGGATGGAGTTCTGGGCTGCGAG 4398

Search completed: January 21, 2003, 09:26:00
Job time : 149.858 secs

```


XX WPI: 2000-256580/22.
DR P-PSDB; AAY70539.
XX
PT Factor 8 homolog polypeptides and nucleic acids encoding them for
PT treating coagulation related disorders such as hemophilia and stroke
XX
PS Claim 1: Page 61-64; 68pp; English.
XX
CC The present sequence is a cDNA encoding human Factor 8 homologue (F8H)
CC which is a coagulation cofactor. The protein is selectively expressed in
CC haematopoietic, heart and reproductive tissues. It has haemostatic and
CC cardioprotective activities. The F8H contains a factor 5/8 signature
CC and is useful as a therapeutic for treating coagulation related diseases
CC such as haemophilia and stroke. The nucleic acid is useful as
CC hybridisation probe and amplification primer for detecting deficiencies
CC in the level of F8H mRNA, for screening F8H gene mutations and for
CC monitoring regulation of gene expression. Fragments of the nucleic acid
CC are also useful as diagnostic probes and primers, and can be used in
CC screening methods such as those using DNA chips. The present sequence is
CC also useful as a target to screen therapeutically useful modulators
CC of the F8H.
XX
SQ Sequence 2046 BP; 615 A; 457 C; 467 G; 507 T; 0 other;
Query Match 89.5%; Score 2041.4; DB 21; Length 2046;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2042; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 238 TGGGAGATCGGTGTAAGATGGGAGAGAGAGTTCGCATCAAAATTTGGTGACATTTGACATT 297
DB 1 TGGGAGATCGGTGTAAGATGGGAGAGAGTTCGCATCAAAATTTGGTGACATTTGACATT 60
QY 298 GAAGATTCGTGATTCGTACATTTAATTACTTGAGAAATTAATGGAATTTGGAGTCAGC 357
DB 61 GAGATTCGTGATTCGTACATTTAATTACTTGAGAAATTAATGGAATTTGGAGTCAGC 120
QY 358 AGAATCGAATAGGCAAAATCTGTGCTGGGTTCCAAATGAACCATTTCAATTTGAATCA 417
DB 121 AGAATCGAATAGGCAAAATCTGTGCTGGGTTCCAAATGAACCATTTCAATTTGAATCA 180
QY 418 AAGGCAATGAATACATCTGTTTCATGAGTGGAAATCCATGTTCTGGAGCGGATTT 477
DB 181 AAGGCAATGAATACATCTGTTTCATGAGTGGAAATCCATGTTCTGGAGCGGATTT 240
QY 478 TTGGCTCATACCTCTGTTATAGATAAACAAGATCTAATTAATCTGTTGGACACTGCATCC 537
DB 241 TTGGCTCATACCTCTGTTATAGATAAACAAGATCTAATTAATCTGTTGGACACTGCATCC 300
QY 538 AATTTTGGAACTGAGTTTCAGTAAGTACTGCCAGCTGGTGTCTGCTTCTCTTTGCT 597
DB 301 AATTTTGGAACTGAGTTTCAGTAAGTACTGCCAGCTGGTGTCTGCTTCTCTTTGCT 360
QY 598 GAGATATCGGAACAATTCCTCATGATATAGATATTCGCAATGTCATGTCATGTCGTGT 657
DB 361 GAGATATCGGAACAATTCCTCATGATATAGATATTCGCAATGTCATGTCATGTCGTGT 420
QY 658 GTGCATCGAGGAGTAGTGTCAACACGTTGGGGCCCAATCAGTCTTCTAATTAATAA 717
DB 421 GTGCATCGAGGAGTAGTGTCAACACGTTGGGGCCCAATCAGTCTTCTAATTAATAA 480
QY 718 GGTATTCCTATTTATGAAAGTTCTTTGGCTAACAACGTCACATCTGCTGGTGACACTTA 777
DB 481 GGTATTCCTATTTATGAAAGTTCTTTGGCTAACAACGTCACATCTGCTGGTGACACTTA 540
QY 778 TCTAAGATCTTTTACATTTAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCT 837
DB 541 TCTAAGATCTTTTACATTTAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCT 600
QY 838 GGTGTGATCGCGATCCTCAATAACAGCATCATCTGCTGGAGTGGACTGACACACA 897
DB 601 GGTGTGATCGCGATCCTCAATAACAGCATCATCTGCTGGAGTGGACTGACACACA 660

QY 898 GGGCAAGAGACAGTTGGAAACCCAAAAGCCAGGCTGAAAACAACTGGACGCGCTTGG 957
DB 661 GGGCAAGAGACAGTTGGAAACCCAAAAGCCAGGCTGAAAACAACTGGACGCGCTTGG 720
QY 958 GGTGCTTTTGGCCACTGATGAATACCAAGTGGTTACAATAGATTGGAATAGGAAAGAAA 1017
DB 721 GGTGCTTTTGGCCACTGATGAATACCAAGTGGTTACAATAGATTGGAATAGGAAAGAAA 780
QY 1018 ATAACAGGCATTATTAACCACTGGATCCACATGGTGGAGCACAATTTACTATGTCTGCC 1077
DB 781 ATAACAGGCATTATTAACCACTGGATCCACATGGTGGAGCACAATTTACTATGTCTGCC 840
QY 1078 TACAGAACTCTGTACAGTGTATGGGAGAAATGAGCTGTGTACAGAGAGCCTGGTGTG 1137
DB 841 TACAGAACTCTGTACAGTGTATGGGAGAAATGAGCTGTGTGTACAGAGAGCCTGGTGTG 900
QY 1138 GAGCAAGATAAGATATTTCAAGSAAACAAGATTATCACAGGATGTGCGTAATAACTTT 1197
DB 901 GAGCAAGATAAGATATTTCAAGSAAACAAGATTATCACAGGATGTGCGTAATAACTTT 960
QY 1198 TTGCCACCAATTTATTCACGCTTTTATTAGAGTGAATCTTACCCAATGGCAGCAGAAAAT 1257
DB 961 TTGCCACCAATTTATTCACGCTTTTATTAGAGTGAATCTTACCCAATGGCAGCAGAAAAT 1020
QY 1258 GCCATGAAATGGAGCTGCTCGATGTCAGTTTATTCCTTAAAGTGTGCTCTCCAAAACCT 1317
DB 1021 GCCATGAAATGGAGCTGCTCGATGTCAGTTTATTCCTTAAAGTGTGCTCTCCAAAACCT 1080
QY 1318 ACTCAACCTCCACCTCCTCGGACACAAATGACCTCAAAACACTACAGCCCTCCAAA 1377
DB 1081 ACTCAACCTCCACCTCCTCGGACACAAATGACCTCAAAACACTACAGCCCTCCAAA 1140
QY 1378 ATAGCCAAAGTCTGCTCCCAAAATTTAGCAACCACTTACAACTCGCAGTAGCAATGAA 1437
DB 1141 ATAGCCAAAGTCTGCTCCCAAAATTTAGCAACCACTTACAACTCGCAGTAGCAATGAA 1200
QY 1438 TTTCTGTGCACAGACAGAACAACTGCCAGTCTGTATATAGAAATACTACCGTAAT 1497
DB 1201 TTTCTGTGCACAGACAGAACAACTGCCAGTCTGTATATAGAAATACTACCGTAAT 1260
QY 1498 CCAATGTAAACAAAGATGAGCGCTGGCTGCTGCTTCTGCTCCCTGCTGTCATGGTC 1557
DB 1261 CCAATGTAAACAAAGATGAGCGCTGGCTGCTGCTTCTGCTCCCTGCTGTCATGGTC 1320
QY 1558 CTCACCTACTCTCATTTCTATATTAGTGTCTTGGCACTGGAGAACACAGAAAGAAA 1617
DB 1321 CTCACCTACTCTCATTTCTATATTAGTGTCTTGGCACTGGAGAACACAGAAAGAAA 1380
QY 1618 ACTGAAGGCACCTATGACTTACCTTACTGGGACCGGGCAGTTGGTGGAAAGGAATGAAG 1677
DB 1381 ACTGAAGGCACCTATGACTTACCTTACTGGGACCGGGCAGTTGGTGGAAAGGAATGAAG 1440
QY 1678 CAGTTTCTTCTGCAAAAGCAGTGGACCATGAGAAACCCAGTTGCTATAGCAGCAGC 1737
DB 1441 CAGTTTCTTCTGCAAAAGCAGTGGACCATGAGAAACCCAGTTGCTATAGCAGCAGC 1500
QY 1738 GAAGTTAATCACTGAGTCCCAAGAGAAAGTCAACACAGTGTGTCAGGCTGACTCTGCAGAG 1797
DB 1501 GAAGTTAATCACTGAGTCCCAAGAGAAAGTCAACACAGTGTGTCAGGCTGACTCTGCAGAG 1560
QY 1798 TATGCTACGCCACTGTGTAGSAGAAATTTGGTGTACACTTTCATCAAAAGATCTACTCTTAAA 1857
DB 1561 TATGCTACGCCACTGTGTAGSAGAAATTTGGTGTACACTTTCATCAAAAGATCTACTCTTAAA 1620
QY 1858 CCAGAAGAGGAAACAGCAGGCTATGAGACCTTAGATCTTACAACTCACCAGGCGAG 1917
DB 1621 CCAGAAGAGGAAACAGCAGGCTATGAGACCTTAGATCTTACAACTCACCAGGCGAG 1680
QY 1918 GAAGTTTATCATGCCCTTATGCTGAACCACTCCCAATTAACGGGCTGAGTATGCAACCCCA 1977
DB 1681 GAAGTTTATCATGCCCTTATGCTGAACCACTCCCAATTAACGGGCTGAGTATGCAACCCCA 1740
QY 1978 ATCATCATGGACATGTGAGGGCACCCCAACAACTTTCAGTTGGTGTGTCAGCCCTCCACATCCACT 2037


```
QY 905 AGACAGTTGGAAACCC-AAAAAGCCAGGCTGAAAAAACCCTGGACCCCTGGGCTGCT 963
Db 930 AGACAGTTGGAAACCCAAAAAGCCAGGCTGAAAAAACCCTGGACCCCTGGG-TGCT 988
QY 964 TTGGCCACTGATGAATACCACTGGTGTACAAATAGATTGAATAAGGAAAGAAATAACA 1023
Db 989 TTGGCCACTGATGAATACCACTGGTGTACAAATAGATTGAATAAGGAAAGAAATAACA 1048
QY 1024 GGCATTATTAACCACTGGATCCACCTGATGGTGGAGACAATTAATATGTCTGCCCTACAGA 1083
Db 1049 GGCATTATTAACCACTGGATCCACCTGATGGTGGAGACAATTAATATGTCTGCCCTACAGA 1107
QY 1084 ATCCCTGACAGTATGATGGGAGAAATGAGTGTACAGAGACCTGGT-GTGGACA 1142
Db 1108 ATCCCTGACAGTATGATGGGAGAAATGAGTGTACAGAGACCTGGTGTGGAGACA 1167
QY 1143 AGATAAGATATTTCAGGAAACAAA--GATTATCACCAAGGAT-GTCCGCTTAATACTTTT 1199
Db 1168 AGATAAGATATTTCAGGAAACAAAAGATTATCACCAAGGATGGTGGCTTAATACTTTT 1227
QY 1200 GGCACCAATTATGACAGTTTATTAGAGTGAATCTTACCCTGAGGACGAGAAATGCG 1259
Db 1228 GGCACCAATTATGACAG-TTTATTAGAGTGAATCTTACCCTGAGGACGAGAAATGCG 1286
QY 1260 CATGAAATGGAGCTGCTCGGATGTCAGTTTATCTCTAAAGTGTGCTCTCCAAAACCTTAC 1319
Db 1287 CATGAAATGGAGCTGCTCGGATGTCAGTTTATCTCTAAAGTGTGCTCTCCAAAACCTTAC 1346
QY 1320 TCAACCTCCACCTCTCTCGGAACGCAATGACCTCAAAAACACTACAGCCCTCCAAAAA- 1378
Db 1347 TCAACCTCCACCTCTCTCGGAACGCAATGACCTC-AAAAACACTACAGCCCTCCAAAAA 1405
QY 1379 --TAGCCAAAGTGTGCTGCCCAAAATTTAGCAACCACTACAACTCGCAGTAGCAATGA 1436
Db 1406 TTTAGCCAAAGTGTGCTGCCCAAAATTTAGCAACCACTACAACTCGCAGTAGCAATGA 1465
QY 1437 ATTTCTGTCACAGACAGAACAACTGCTGCTGATATCAGAAATACCTGCTTAC 1496
Db 1466 ATTTCTGTCACAGACAGAACAACTGCTGCTGATATCAGAAATACCTGCTTAC 1525
QY 1497 TCCAAATGAACCAAGATGTAGCGCTGCTGCTGCTTGTTCCTGCTGCTGCTGCTGCTGCT 1556
Db 1526 TCCAAATGAACCAAGATGTAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1585
QY 1557 CCTCACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1616
Db 1586 CCTCACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1645
QY 1617 AACTGAAGCACCCTATGACTTACCTTACCTGAGGACCGGCGAGGT 1659
Db 1646 AACTGAAGCACCCTATGACTTACCTTACCTGAGGACCGGCGAGGT 1688
```

RESULT 6

ABK83640

ID ABK83640 standard; cDNA; 1388 BP.

XX AC ABK83640;

XX DT

XX 14-AUG-2002 (first entry)

DE Human cDNA differentially expressed in granulocytic cells #211.

XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic inflammation; protozoal infection;
KW fungal infection; sterile inflammation; disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammation; bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.

XX W0200228999-A2.

XX 11-APR-2002.

XX 03-OCT-2001; 2001WO-US30821.

XX 03-OCT-2000; 2000US-237189P.

XX (GENE-) GENE LOGIC INC.

PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

XX WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression
of genes associated with granulocyte activation, which serves as
diagnostic markers that is useful for monitoring disease states and
drug toxicity -

XX Claim 1; SEQ ID No 211; 114pp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation
(GCA), by detecting the level of expression of gene(s) (Gs) identified by
DNA chip analysis as given in the specification, and comparing
the expression level to an expression level in an unactivated
GC, where differential expression of Gs is indicative of GCA.
XX Also included are modulating (M2) GA by contacting GC with an agent
that alters the expression of at least one gene in Gs; (2) screening (M3)
chronic) in a tissue, an allergic response in a subject, exposure of a
subject to a pathogen or sterile inflammatory disease using the
gene expression profile; (3) detecting (M4) an inflammation (especially
chronic) in a tissue, an allergic response in a subject, exposure of a
subject to a pathogen or sterile inflammatory disease, by detecting the
level of expression in a sample of the tissue of gene(s) from Gs, where
the level of expression of the gene is indicative of inflammation;
XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,
an allergic response in a subject, exposure of a subject to a pathogen
or sterile inflammatory disease, by contacting a tissue having
inflammation with an agent that modulates the expression of gene(s)
from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
modulating GA; M3 is useful for screening an agent capable of modulating
GCA preferably in an inflammation in a tissue; M4 is useful for
detecting an inflammation (especially chronic) in a tissue, an allergic
response in a subject, exposure of a subject to a pathogen or sterile
inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
glomerulonephritis, asthma, thrombosis, cardiac reperfusion syndrome,
reflexion injury, ARDS, adult respiratory distress syndrome,
inflammatory bowel disease, Crohn's disease, ulcerative colitis,
periodontal disease; also bacterial infection, viral infection, and
parasitic infection, protozoal infection, fungal infection, and M5 is
useful for treating one of the above conditions. The present
sequence represents a gene differentially expressed in granulocytes.
XX Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic
format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1388 BP; 432 A; 289 C; 302 G; 364 T; 1 other;

Query Match 50.8%; Score 1158.4; DB 24; Length 1388;
Best Local Similarity 97.5%; Pred. No. 0;

Matches 1358; Conservative 0; Mismatches 12; Indels 23; Gaps 17;

QY 143 GTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAACTTACATCACTAACT 202

Db 1 GTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAACTTACATCACTAACT 60

QY 203 ACCCAGACACCTATCCCAACAGCAGTGTGTGAATGGGAGATCCGTGTAAGATGGGAG 262

Db 61 ACCCAGACACCTATCCCAACAGCAGTGTGTGAATGGGAGATCCGTGTAAGATGGGAG 120

Qy	1331	CTCTCGGAACACGATGACCTCAAAAACACATACAGCGCCCTCCAAAAG	1387
Db	1197	CTCTCGGAACACGATGACCTC-AAAACACATACAGCGCCCTCCAAAATTTAGCCAAAG	1255
Qy	1388	GTCTGTGCCCCAAAATTTAGCAACCACTACAACCTCGCAGTAGCAATGAATTTCTCTGCAC	1447
Db	1256	GTCTGTGCCCCAAAATTTAGCAACCACTACAACCTCGCAGTAGCAATGAATTTCTCTGCAC	1315
Qy	1448	AGACAGAACAAACAACCTGCCAGTCTCTGATATCAGAAATACTA--CCGTAACCTCCAAATG	1504
Db	1316	AGACAGAACAAACAACCTGCCAGTCTCTGATATCAGAAATACTACCTACCGTAAACCTCCAAATG	1375
Qy	1505	TAAACCAAGATGT	1517
Db	1376	TAAACCAAGATGT	1388
RESULT 7			
AA565485			
ID	AA565485	standard; cDNA; 594 BP.	
XX	AC		
XX	AA565485;		
XX	AC		
DT	13-FEB-2002	(first entry)	
XX		DNA encoding novel human diagnostic protein #1289.	
DE			
XX		Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW		food supplement; medical imaging; diagnostic; genetic disorder; ss.	
KW			
XX			
OS		Homo sapiens.	
XX			
PN	W0200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
XX	30-MAR-2001; 2001WO-0508631.		
PF			
XX	31-MAR-2000; 2000US-0540217.		
PR			
PR	23-AUG-2000; 2000US-0649167.		
XX			
PR	(HYSE-) HYSEQ INC.		
PA			
XX			
FI	Drmanac RT, Liu C, Tang YT;		
PI			
XX			
DR	WPI: 2001-639362/73.		
DR	P-PSDB; ABG01298.		
XX			
PT		New isolated polynucleotide and encoded polypeptides, useful in	
PT		diagnostics, forensics, gene mapping, identification of mutations	
PT		responsible for genetic disorders or other traits and to assess	
PT		biodiversity	
XX			
PS	Claim 1; SEQ ID No 1289; 103pp; English.		
XX			
CC		The invention relates to isolated polynucleotide (I) and	
CC		polypeptide (II) sequences. (I) is useful as hybridisation probes,	
CC		polymerase chain reaction (PCR) primers, oligomers, and for chromosome	
CC		and gene mapping, and in recombinant production of (II). The	
CC		polynucleotides are also used in diagnostics as expressed sequence tags	
CC		for identifying expressed genes. (I) is useful in gene therapy techniques	
CC		to restore normal activity of (II) or to treat disease states involving	
CC		(II). (II) is useful for generating antibodies against it, detecting or	
CC		quantitating a polypeptide in tissue, as molecular weight markers and as	
CC		a food supplement. (II) and its binding partners are useful in medical	
CC		imaging of sites expressing (II). (I) and (II) are useful for treating	
CC		disorders involving aberrant protein expression or biological activity.	
CC		The polypeptide and polynucleotide sequences have applications in	
CC		diagnostics, forensics, gene mapping, identification of mutations	
CC		responsible for genetic disorders or other traits to assess biodiversity	
CC		and to produce other types of data and products dependent on DNA and	
CC		amino acid sequences. AA564197-AA594564 represent novel human	
CC		diagnostic coding sequences of the invention.	

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 594 BP; 175 A; 160 C; 139 G; 120 T; 0 other;

Query Match 26.1%; Score 594; DB 23; Length 594;

Best Local Similarity 100.0%; Pred. No. 3.5e-167;

Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1672 ATGAAGCAGTTCTTCTCTCAAAAGCAGTGGACCATGAGAAACCCAGTTCGCTATAGC 1731

Db 1 ATGAAGCAGTTCTTCTCTCAAAAGCAGTGGACCATGAGAAACCCAGTTCGCTATAGC 60

QY 1732 AGCAGCGAAGTTAATCACTGAGTCAAGAGAGTCAACACAGTGTGCGAGGCTGACTCT 1791

Db 61 AGCAGCGAAGTTAATCACTGAGTCAAGAGAGTCAACACAGTGTGCGAGGCTGACTCT 120

QY 1792 GCAGATATGCTCAGCCACTGTGAGGAGAAATTTGGTACACTTCATCAAAAGATCTACC 1851

Db 121 GCAGATATGCTCAGCCACTGTGAGGAGAAATTTGGTACACTTCATCAAAAGATCTACC 180

QY 1852 TTTAAACAGAGAAGAAAGAGCAGGCTATGACAGCTAGATCCTTACAACTCACCA 1911

Db 181 TTTAAACAGAGAAGAAAGAGCAGGCTATGACAGCTAGATCCTTACAACTCACCA 240

QY 1912 GGCAGGAAGTTTATCATGCTATGCTGAACCACTCCCAATACGGGGCTGAGTATGCA 1971

Db 241 GGCAGGAAGTTTATCATGCTATGCTGAACCACTCCCAATACGGGGCTGAGTATGCA 300

QY 1972 ACCCAATCATATGACATGTGAGGAGACCCCAACTTCAGTTGGTGCAGCCCTCCACA 2031

Db 301 ACCCAATCATATGACATGTGAGGAGACCCCAACTTCAGTTGGTGCAGCCCTCCACA 360

QY 2032 TCCACITTTCAAGCTACGGGGAACCAACTCCCGCACTAGTGGGAATCTACATACACTT 2091

Db 361 TCCACITTTCAAGCTACGGGGAACCAACTCCCGCACTAGTGGGAATCTACATACACTT 420

QY 2092 CTCTCCAGGACTGACAGTCTCTCTCAGCCAGCCAGCTATGATACCCGGAAGCTGGG 2151

Db 421 CTCTCCAGGACTGACAGTCTCTCTCAGCCAGCCAGCTATGATACCCGGAAGCTGGG 480

QY 2152 AAGCCAGGCTACCTGCCCGAGACGAATTTGGTGTACAGGTGCCAGAGCACACAAGAA 2211

Db 481 AAGCCAGGCTACCTGCCCGAGACGAATTTGGTGTACAGGTGCCAGAGCACACAAGAA 540

QY 2212 GTATCAGGACAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGA 2265

Db 541 GTATCAGGACAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGA 594

RESULT 8

AZ80396/C

ID AZ80396 standard; cDNA; 640 BP.

XX AC

XX AZ80396;

DT 07-APR-2000 (first entry)

XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:480.

DE Human; gene expression product; diagnosis; tumor; colon cancer;

XX colorectal adenocarcinoma; cell line SW480; cell proliferation;

KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;

KW hyperplasia; ds.

XX Homo sapiens.

OS WO9964576-A2.

XX 16-DEC-1999.

XX 09-JUN-1999; 99WO-IB01062.

XX

XX

XX

XX

XX 10-JUN-1998; 98US-0088801.

XX (FARB) BAYER CORP.

XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;

PI Carroll E, Catino JV, Derti A, Ford DW, Lewis ME, Monahan JE;

PI Schlegel R;

XX WPI; 2000-087220/07.

DR

XX Novel nucleic acids, used to develop products for the diagnosis and

PT treatment of disorders involving unwanted cell proliferation,

PT particularly cancers, especially colon cancer

XX Claim 15; Page 322; 469pp; English.

XX AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from

CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The

CC cDNA clones can be used to generate antisense oligonucleotides which

CC can be used for antisense therapy. Methods and products from the present

CC invention can be used for identifying and/or classifying cancerous cells

CC present in a human tumour, particularly in solid tumours, e.g.

CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones

CC can be used for developing agents for the diagnosis and treatment of

CC disorders involving unwanted cell proliferation, such as neoplasia,

CC dysplasia or hyperplasia.

XX Sequence 640 BP; 148 A; 134 C; 160 G; 192 T; 6 other;

SQ

Query Match 23.9%; Score 544.6; DB 21; Length 640;

Best Local Similarity 96.7%; Pred. No. 2.4e-152;

Matches 616; Conservative 0; Mismatches 15; Indels 6; Gaps 6;

QY 1196 TTTTGGCCACCAATATTGACGCTTTTATTAGAGTGAATCTCTACCATGCGACGAGAAA 1255

Db 632 TTTTGGCCACCAATATTGACGCTTTTATTAGAGTGAATCTCTACCATGCGACGAGAAA 576

QY 1256 TTGGCCATGAAATGGAGCTGCTCGGATGTCAGTTTATTCCTAAGGTGCTCTCCAAAAC 1315

Db 575 TTGGCCATGAAAT-GAAGTCTCGGATGTCAGTTTATTCCTAAGGTGCTCTCCAAAAC 518

QY 1316 TTACTCAACTCCACCTCTCGGAGCAGCAATGACCTCAAAAACACTACAGCCCTCCAA 1375

Db 517 TTACTCAACTCCACCTCTCTGNAACAGCAATGCTCAAAAACACTACAGCCCTCC-A 459

QY 1376 AAATAGCCAAAGGTCGTGCCCCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATG 1435

Db 458 AAATAGCCAAAGGTCGTGCCCCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATG 399

QY 1436 AATTTCTCTCAGACAGACAAACAACTGCCAGTCTGTATATCAGAAATACCTACCTAA 1495

Db 398 AATTTCTCTCAGACAGACAAACAACTGCCAGTCTGTATATCAGAAATACCTACCTAA 339

QY 1496 CTCCAAATGTAACCAAGATGTAGCGCTGGCTCAGTTCTTGTCCCTGTGCTGGTCTAGG 1555

Db 338 CTCCAAATGTAACCAAGATGTAGCGCTGGCTCAGTTCTTGTCCCTGTGCTGGTCTAGG 279

QY 1556 TCCTCACTACTCTCATCTCTCATATTAGTGTGCTGGCACTGGAGAAACAGAGAA 1615

Db 278 TCCTCACTACTCTCATCTCTCATATTAGTGTGCTGGCACTGGAGAAACAGAGAA 219

QY 1616 AAAGTGAAGGACCTATGACTTACCTTACTGGGACGGGAGGTTGGTGGAAAGGAATGA 1675

Db 218 AAAGTGAAGGACCTATGACTTACCTTACTGGGACGGGAGGTTGGTGGAAAGGAATGA 159

QY 1676 AGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAAACCCAGTTTCGTATAGCAGCA 1735

Db 158 AGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAAACCCAGTTTCGTATAGCAGCA 99

QY 1736 GCGAAGTTAATCACTCTGAGTCCCAAGAGAGTCAACCACTGCTGAGGCTGACTCTGCAG 1795

Db 98 GCGAAGTTAATCACTCTGAGTCCCAAGAGAGTCAACCACTGCTGAGGCTGACTCTGCAG 39

QY 1796 AGTATGCTCAGCCACTGGTAGGAGGAATTTGGGTAC 1832
 |||||
 Db 38 ACTATGCTCAGCCACTGGTAGGAGGAATTTGGGTAC 2

RESULT 9
 ABA60693
 ID ABA60693 standard; DNA; 580 BP.

XX AC ABA60693;

XX DT 01-FEB-2002 (first entry)

XX DE Human foetal liver single exon nucleic acid probe #8998.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX OS Homo sapiens.

XX PN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00669.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX Claim 1; SEQ ID NO 8998; 639pp + sequence listing; English.

XX CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.

XX CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;

Query Match 21.4%; Score 488.8; DB 22; Length 580;

Best Local Similarity 99.6%; Pred. NO. 1.2e-135;

Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1789 TCTGTCAGATGCTCAGCCACTGGTAGGAGGAATTTGGGTACATCAAGATCT 1848
 |||||

Db 64 TCCACAGATGCTCAGCCACTGGTAGGAGGAATTTGGGTACATCAAGATCT 123

QY 1849 ACCTTTAAACCAAGAGGAAGAAAGACGAGGCTATCAGACCTAGATCTTACAATCA 1908
 |||||

Db 124 ACCTTTAAACCAAGAGGAAGAAAGACGAGGCTATCAGACCTAGATCTTACAATCA 183

QY 1909 CCAGGGCAGGAAGTTTATCATGCTTATGCTGTAACCATCCCAATTAGGGCCCTGAGTAT 1968
 |||||

Db 184 CCAGGGCAGGAAGTTTATCATGCTTATGCTGTAACCATCCCAATTAGGGCCCTGAGTAT 243

QY 1969 GCAACCCCAATCATCATGACATGTGAGGACCCCAACCTTCACTTGGTCAGCCCTCC 2028
 |||||
 Db 244 GCAACCCCAATCATCATGACATGTGAGGACCCCAACCTTCACTTGGTCAGCCCTCC 303
 |||||
 QY 2029 ACATCCACTTTCAAGGCTACGGGGAACCAACCTTCCCCCACTAGTGGGAACTTACAATACA 2088
 |||||
 Db 304 ACATCCACTTTCAAGGCTACGGGGAACCAACCTTCCCCCACTAGTGGGAACTTACAATACA 363
 |||||
 QY 2089 CTTTCTCTCCAGGACTGACAGCTGCTCTCTCAGCCCAAGCCCACTATGATATACCCGAAAGCT 2148
 |||||
 Db 364 CTTTCTCTCCAGGACTGACAGCTGCTCTCTCAGCCCAAGCCCACTATGATATACCCGAAAGCT 423
 |||||
 QY 2149 GGGAGCCAGGCTTACCTGCCCCAGACAGAAATGGGTGTACAGGTGCCACAGACACAA 2208
 |||||
 Db 424 GGGAGCCAGGCTTACCTGCCCCAGACAGAAATGGGTGTACAGGTGCCACAGACACAA 483
 |||||
 QY 2209 GAAGTATCAGGAGCAGAGGAGGATGGGAATGTGATGTTTTAAAGAAATCCTTTGAAGA 2268
 |||||
 Db 484 GAAGTATCAGGAGCAGAGGAGGATGGGAATGTGATGTTTTAAAGAAATCCTTTGAAGA 543
 |||||

RESULT 10

AAK08975

ID AAK08975 standard; DNA; 580 BP.

XX AC AAK08975;

XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe SEQ ID NO: 8966.

XX KW Human; brain expressed exon; gene expression analysis; probe;

XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX KW epilepsy; cancer; ss.

XX OS Homo sapiens.

XX PN WO200157275-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00667.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483446/52.

XX PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX Example 4; SEQ ID NO: 8966; 650pp + Sequence Listing; English.
 XX PS The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.

XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 25; SEQ ID No 9268; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX CC
XX SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;

Query Match 21.4%; Score 488.8; DB 22; Length 580;
Best Local Similarity 99.6%; Pred. No. 1.2e-135;
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1789 TCCTGACAGTATCTCAGCCACGTAGGAGGAATTTGGTACACTTCATCAAGATCT 1848
DB 64 TCACAGAGTATCTCAGCCACGTGGTAGGAGGAATTTGGTACACTTCATCAAGATCT 123

QY 1849 ACCTTTAAACAGAGAAGAAAGAGGAGGCTATGCAGACCTAGATCCTTACAACCTCA 1908
DB 124 ACCTTTAAACAGAGAAGAAAGAGGAGGCTATGCAGACCTAGATCCTTACAACCTCA 183

QY 1909 CCAGGGCAGGAAGTTTATCATGCTTATGCTGAACCACTCCCAATTACGGGGCCTGAGTAT 1968
DB 184 CCAGGGCAGGAAGTTTATCATGCTTATGCTGAACCACTCCCAATTACGGGGCCTGAGTAT 243

QY 1969 GCAACCCCAATCATATGACATGTGACGATGTGAGGCACCCCACTTCACTTGTGTCAGCCCTCC 2028
DB 244 GCAACCCCAATCATATGACATGTGACGATGTGAGGCACCCCACTTCACTTGTGTCAGCCCTCC 303

QY 2029 ACATCCACTTTCAAGCTACGGGGAACCACTCCCACTAGTGGGAATTTACAATACA 2088
DB 304 ACATCCACTTTCAAGCTACGGGGAACCACTCCCACTAGTGGGAATTTACAATACA 363

QY 2089 CTTCTCTCCAGGACTGACAGCTGCTCCTCAGCCAGCCAGCCAGTATGATACCCGAAAGCT 2148
DB 364 CTTCTCTCCAGGACTGACAGCTGCTCCTCAGCCAGCCAGCCAGTATGATACCCGAAAGCT 423

QY 2149 GGAAGCCAGGCTTACTGCCCCAGAGGAATTTGGTGTACAGGTGCCCCAGAGACACAA 2208
DB 424 GGAAGCCAGGCTTACTGCCCCAGAGGAATTTGGTGTACAGGTGCCCCAGAGACACAA 483

QY 2209 GAAGTATCAGGACGAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGAGA 2268
DB 484 GAAGTATCAGGACGAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGAGA 543

QY 2269 TGATGCTGCTTT 2280

DB 544 TGATGCTGCTTT 555

RESULT 13
ABA76819
ID ABA76819 standard; DNA; 467 BP.
XX AC ABA76819;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #25124.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PS Claim 4; SEQ ID NO 25124; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC foetal liver. The present sequence is a single exon nucleic acid
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;

Query Match 20.5%; Score 467; DB 22; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.5e-129;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1796 AGTATGCTCAGCCACTGGTAGGAGGAATTTGGTACACTTCATCAAGATCTACCTTTA 1855
DB 1 AGTATGCTCAGCCACTGGTAGGAGGAATTTGGTACACTTCATCAAGATCTACCTTTA 60

QY 1856 AACCAAGAAGGAAGAAAGAGCAGGCTATGCAGACCTAGATCCTTCAACTCACCAGGC 1915
DB 61 AACCAAGAAGGAAGAAAGAGCAGGCTATGCAGACCTAGATCCTTCAACTCACCAGGC 120

QY 1916 AGGAAGTTTATCATGCTGCTGAACCACTCCCAATTACGGGGCCTGAGTATGCAACCC 1975
DB 121 AGGAAGTTTATCATGCTGCTGCTGAACCACTCCCAATTACGGGGCCTGAGTATGCAACCC 180

QY 1976 CAATCATCATGAGCATGTCAGGGCACCCCAACTTCAGTTGCTGACCCCTCCACATCCA 2035

Db 181 CAATCATCATGACATGTCTCAGGCGACCCACCAACTTCAGTTGGTCAGCCCTCCACATCCA 240
QY 2036 CTTTCAAGCTACGGGGAAACCAACCTCCCTCCTAGTGGGAACCTTACATACACTTCTCT 2095
Db 241 CTTTCAAGCTACGGGGAAACCAACCTCCCTCCTAGTGGGAACCTTACATACACTTCTCT 300
QY 2096 CCAGGACTCACAGCTGCTCCTCAGCCAGGCGCCAGTATGATACCCCGAAAGCTGGGAAGC 2155
Db 301 CCAGGACTCACAGCTGCTCCTCAGCCAGGCGCCAGTATGATACCCCGAAAGCTGGGAAGC 360
QY 2156 CAGGCTACTGCCCCAGACGAATTGGTGTACAGGTGCCAGAGACACACAGAAGTAT 2215
Db 361 CAGGCTACTGCCCCAGACGAATTGGTGTACAGGTGCCAGAGACACACAGAAGTAT 420
QY 2216 CAGGAGCAGGAGGATGGGGAATGTGATGTTTAAAGAAATCCTT 2262
Db 421 CAGGAGCAGGAGGATGGGGAATGTGATGTTTAAAGAAATCCTT 467
RESULT 14
AAK25439
ID AAK25439 standard; DNA; 467 BP.
XX AC AAK25439;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe SEQ ID NO: 25430.
XX KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX PS Example 4; SEQ ID NO: 25430; 650pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;
Query Match 20.5%; Score 467; DB 22; Length 467;
Best Local Similarity 100.0%; Pred. No. 3.5e-129; Indels 0; Gaps 0;
Matches 467; Conservative 0; Mismatches 0;

QY 1796 AGTATGCTCAGCCTGGTAGGAGGAATGTTGGTACACTTTCATCAAAAGATCTACCTTTA 1855
Db 1 AGTATGCTCAGCCTGGTAGGAGGAATGTTGGTACACTTTCATCAAAAGATCTACCTTTA 60
QY 1856 AACCAAGAAGAAAGAACGAGGCTATGCAGACCTAGATCCTTACAACTCACCAGGCG 1915
Db 61 AACCAAGAAGAAAGAACGAGGCTATGCAGACCTAGATCCTTACAACTCACCAGGCG 120
QY 1916 AGGAAGTTTATCATGCTATGCTGAACACTCCTCCAAATTACGGGGCTGAGTATGCAACCC 1975
Db 121 AGGAAGTTTATCATGCTATGCTGAACACTCCTCCAAATTACGGGGCTGAGTATGCAACCC 180
QY 1976 CAATCATCATGACATGTCTCAGGCGACCCACAACTTTCAGTTGGTCAGCCCTCCACATCCA 2035
Db 181 CAATCATCATGACATGTCTCAGGCGACCCACAACTTTCAGTTGGTCAGCCCTCCACATCCA 240
QY 2036 CTTTCAAGGCTACGGGGAACCAACCTCCCTCCTAGTGGGAACCTTACAACTACACTTCTCT 2095
Db 241 CTTTCAAGGCTACGGGGAACCAACCTCCCTCCTAGTGGGAACCTTACAACTACACTTCTCT 300
QY 2096 CCAGGACTCACAGCTGCTCCTCAGCCAGGCGCCAGTATGATACCCCGAAAGCTGGGAAGC 2155
Db 301 CCAGGACTCACAGCTGCTCCTCAGCCAGGCGCCAGTATGATACCCCGAAAGCTGGGAAGC 360
QY 2156 CAGGCTACTGCCCCAGACGAATTGGTGTACAGGTGCCAGAGACACACAGAAGTAT 2215
Db 361 CAGGCTACTGCCCCAGACGAATTGGTGTACAGGTGCCAGAGACACACAGAAGTAT 420
QY 2216 CAGGAGCAGGAGGATGGGGAATGTGATGTTTAAAGAAATCCTT 2262
Db 421 CAGGAGCAGGAGGATGGGGAATGTGATGTTTAAAGAAATCCTT 467

RESULT 15
AAK51451
ID AAK51451 standard; DNA; 467 BP.
XX AC AAK51451;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 26008.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 26008; 658pp + Sequence Listing; English.

XX

CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.

XX

SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;

Query Match 20.5%; Score 467; DB 22; Length 467;

Best Local Similarity 100.0%; Pred. No. 3.5e-129;

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1796 AGTATGCTCAGGCACCTGGTAGGAGGAATGTTGGTACACTTCATCAAGATCTACCTTTA 1855

DB 1 AGTATGCTCAGGCACCTGGTAGGAGGAATGTTGGTACACTTCATCAAGATCTACCTTTA 60

QY 1856 AACCCAGAAGAAAGAACAGCAGGCTATGCAGACCTAGATCCTTACAACCTACCCAGGGC 1915

DB 61 AACCCAGAAGAAAGAACAGCAGGCTATGCAGACCTAGATCCTTACAACCTACCCAGGGC 120

QY 1916 AGGAAGTTTATCATGCTGATGTGAACCACTCCCAATTACGGGGCCTGAGTATGCAACC 1975

DB 121 AGGAAGTTTATCATGCTGATGTGAACCACTCCCAATTACGGGGCCTGAGTATGCAACC 180

QY 1976 CAATCATATGGACATGTCAGGSCACCCCACTTCAGTTGGTCAGCCCTCCACATCCA 2035

DB 181 CAATCATATGGACATGTCAGGSCACCCCACTTCAGTTGGTCAGCCCTCCACATCCA 240

QY 2036 CTTTCAAGGCTACGGGAACCACTCCCACTAGTGGGAACCTTACAATACACTTCTCT 2095

DB 241 CTTTCAAGGCTACGGGAACCACTCCCACTAGTGGGAACCTTACAATACACTTCTCT 300

QY 2096 CCAGGACTGACAGCTGCTCCTCAGCCAGCCAGTATGATATACCCGAAAGCTGGGAAGC 2155

DB 301 CCAGGACTGACAGCTGCTCCTCAGCCAGCCAGTATGATATACCCGAAAGCTGGGAAGC 360

QY 2156 CAGGTCTACCTGCCCCAGACGAATTGGTGTACAGGTGCCACAGAGCACACAGAGATAT 2215

DB 361 CAGGTCTACCTGCCCCAGACGAATTGGTGTACAGGTGCCACAGAGCACACAGAGATAT 420

QY 2216 CAGGAGCAGGAGGGATGGGAATGTGATCTTTTAAAGAAATCCTT 2262

DB 421 CAGGAGCAGGAGGGATGGGAATGTGATCTTTTAAAGAAATCCTT 467

Search completed: January 21, 2003, 02:15:32

Job time : 475.72 secs


```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5528102"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT      312 a 280 c 236 g 237 t
ORIGIN
Query Match      36.8%; Score 838; DB 13; Length 1065;
Best Local Similarity 97.0%; Pred. No. 7.1e-215;
Matches 865; Conservative 0; Mismatches 25; Indels 2; Gaps 1;

QY 1390 CTTGCCCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATGAAATTTCTCTGCACAG 1449
      |||||||
Db 1 CTTGCCCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATGAAATTTCTCTGCACAG 60

QY 1450 ACAGACAAACACTGCCAGTCTGTATATCAGAAATACCTACGTAACCTCCAAATTAACC 1509
      |||||||
Db 61 ACAGACAAACACTGCCAGTCTGTATATCAGAAATACCTACGTAACCTCCAAATTAACC 120

QY 1510 AAAGATGTAGCGCTGGCTGCAGTCTTGTCCCTGTGCTGGTTCATGCTCCTACTACTCTC 1569
      |||||||
Db 121 AAAGATGTAGCGCTGGCTGCAGTCTTGTCCCTGTGCTGGTTCATGCTCCTACTACTCTC 180

QY 1570 ATTCTCATATTAGTGTGCTTTGGCACTGGGAAACAGAAAGAAAAAACTGAAGGCACC 1629
      |||||||
Db 181 ATTCTCATATTAGTGTGCTTTGGCACTGGGAAACAGAAAGAAAAAACTGAAGGCACC 240

QY 1630 TATGACTTACCTTACTGGGACCGGCGAGTGTGGTGAAGGAATGAAGCAGTTTCTTCT 1689
      |||||||
Db 241 TATGACTTACCTTACTGGGACCGGCGAGTGTGGTGAAGGAATGAAGCAGTTTCTTCT 300

QY 1690 GCAAAAGCAGTGGACCATGAGGAACCCAGTTCGCTATAGCAGGAGCGGAAGTTTATAC 1749
      |||||||
Db 301 GCAAAAGCAGTGGACCATGAGGAACCCAGTTCGCTATAGCAGGAGCGGAAGTTTATAC 360

QY 1750 CTGAGTCCAGAGAACTACCAAGTGTGTCAGGCTGACTCTGCAAGATATGCTCAGCCA 1809
      |||||||
Db 361 CTGAGTCCAGAGAACTACCAAGTGTGTCAGGCTGACTCTGCAAGATATGCTCAGCCA 420

QY 1810 CTGGTAGGAGGAATTTGGTACACTTCATCAAGATCTACCTTTAAACCAGGAAGCA 1869
      |||||||
Db 421 CTGGTAGGAGGAATTTGGTACACTTCATCAAGATCTACCTTTAAACCAGGAAGCA 480

QY 1870 AAAGAACGAGGCTATGACAGCTTAGATCCTTACAACTACCAAGGCGAGGAAGTTTATCAT 1929
      |||||||
Db 481 AAAGAACGAGGCTATGACAGCTTAGATCCTTACAACTACCAAGGCGAGGAAGTTTATCAT 540

QY 1930 GCTATGCTGAACACTTCCCAATTACGGGGCCCTGAGTATGCAACCCCAATCATATGAGAC 1989
      |||||||
Db 541 GCTATGCTGAACACTTCCCAATTACGGGGCCCTGAGTATGCAACCCCAATCATATGAGAC 600

QY 1990 ATGTCAGGCAACCCACAACTTCAGTTGTCAGCCCTCCACATCCACTTTCAGGCTAGC 2049
      |||||||
Db 601 ATGTCAGGCAACCCACAACTTCAGTTGTCAGCCCTCCACATCCACTTTCAGGCTAGC 660

QY 2050 GGGAAACCAACTCCCTCCACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGC 2109
      |||||||
Db 661 GGGAAACCAACTCCCTCCACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGC 720

QY 2110 TGCTCTCAGCCCGCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGCTCTACCTGCC 2169
      |||||||
Db 721 TGCTCTCAGCCCGCCAGTATGATACCCCGAAAGCTGGGAAGCCAGGCTCTACCTGCC 780

QY 2170 CCAGACGAATTTGTTACCAAGTGTCCACAGAGCACACAGAAAGTATCAGGAGCAGGAAG 2229
      |||||||
Db 781 CCAGACGAATTTGTTACCAAGTGTCCACAGAGCACACAGAAAGTATCAGGAGCAGGAAG 840
```

```
QY 2230 G--ATGGGGAATGTGATGTTTAAAGAAATCCTTTGAAGATGATGCTGCTT 2279
      |||||||
Db 841 GGATCGGGAATGTGATGTTTAAAGAAATCCTTTGAAGAAATGATGCTT 892

RESULT 2
BM905706      1061 bp mRNA linear EST 12-MAR-2002
LOCUS        AGNCOURT_6721242 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5556057
DEFINITION   5', mRNA sequence.
ACCESSION    BM905706
VERSION       BM905706.1 GI:19356085
KEYWORDS     EST.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE    1 (bases 1 to 1061)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgaaps-remail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM12277 row: d column: 10
              High quality sequence stop: 629.

FEATURES             Location/Qualifiers
     source            1..1061
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5556057"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;
                     Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                     Average insert size 2.1 kb."
BASE COUNT          281 a 241 c 255 g 283 t 1 others
ORIGIN
Query Match        33.3%; Score 759; DB 14; Length 1061;
Best Local Similarity 94.1%; Pred. No. 1.6e-193;
Matches 867; Conservative 0; Mismatches 40; Indels 14; Gaps 7;

QY 37 CGCTCCCTCCCTCCCTCCCAACTCCTCCTCTCTCCATGCTCTGTTCCTCCTGCTC 96
      |||||||
Db 2 CGCTCCCTCCCTCCCTGCT-CAACTCCTCCTCTCTCTCCATGCTCTGTTCCTCCTGCTC 60

QY 97 TTACTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 156
      |||||||
Db 61 TTACTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120

QY 157 CACACTGTACTAGGCCCTCAGAGTGGAGCCCTTACATCCATAAATACCCACAGACCTAT 216
      |||||||
Db 121 CACACTGTACTAGGCCCTCAGAGTGGAGCCCTTACATCCATAAATACCCACAGACCTAT 180

QY 217 CCCACAGACACTGTTTGTGAATGGGAGATCCGTGTAAGATGGAGAGAGATTCGCATC 276
      |||||||
Db 181 CCCACAGACACTGTTTGTGAATGGGAGATCCGTGTAAGATGGAGAGAGATTCGCATC 240

QY 277 AAATTTGGTGACTTTTGCATTTGAAGATTTCTGATTTCTTGTCTCATTCTTACTTTGAGAATT 336
      |||||||
Db 241 AAATTTGGTGACTTTTGCATTTGAAGATTTCTGATTTCTTGTCTCATTCTTACTTTGAGAATT 300

QY 337 TATAATGGAATGGAGTTCAGCAGAACTGAAATAGGCAAAATCTGTGCTGGGGTTGCAA 396
      |||||||
Db 301 TATAATGGAATGGAGTTCAGCAGAACTGAAATAGGCAAAATCTGTGCTGGGGTTGCAA 360
```

source

1..693

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3434806"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrotheoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: ovary; Vector: p773D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGGACATTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT 162 a 173 c 159 g 197 t 2 others
ORIGIN

Query Match 30.1%; Score 687; DB 12; Length 693;
Best Local Similarity 99.7%; Pred. No. 3.6e-174;
Matches 687; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 GCCGCGCGCGCGCTGGGCGCGCTCCCTCTCCGCTCCCTCCCTCCCTCCCTCCCAAC 60
DB 4 GCCGCGCGCGCGCTGGGCGCGCTCCCTCTCCGCTCCCTCCCTCCCTCCCTCCCAAC 63
QY 61 TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
DB 64 TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 123
QY 121 GAGGAGCTGGAGCGCCAGCAAGGTGATGTGGACACACTGTACTAGCCCTGAGAGT 180
DB 124 GAGGAGCTGGAGCGCCAGCAAGGTGATGTGGACACACTGTACTAGCCCTGAGAGT 183
QY 181 GGAACCTTACATCCATAAATACCCACACACCTATCCCAACAGCACTGTTTGTGATGG 240
DB 184 GGAACCTTACATCCATAAATACCCACACACCTATCCCAACAGCACTGTTTGTGATGG 243
QY 241 GAGATCGTGTAAAGATGGGAGAGAGTTCGCATCAAAATTTGGTGACTTTGACATTTGAA 300
DB 244 GAGATCGTGTAAAGATGGGAGAGAGTTCGCATCAAAATTTGGTGACTTTGACATTTGAA 303
QY 301 GATTCTGATTCTTGTCTCACTTTAATTAATGAGAAATTTAATGAAATGGAGTCAGCAGA 360
DB 304 GATTCTGATTCTTGTCTCACTTTAATTAATGAGAAATTTAATGAAATGGAGTCAGCAGA 363
QY 361 ACTGAATAGGCAAAATCTGTGTCTGGGTTCGAAATGAACCAATTCATTAATGAATCAAAA 420
DB 364 ACTGAATAGGCAAAATCTGTGTCTGGGTTCGAAATGAACCAATTCATTAATGAATCAAAA 423
QY 421 GGCAATGAAATCACATTGCTTTCATGAGTGGAAATCCATGTTCTGGACGCGGATTTTGTG 480
DB 424 GGCATGAAATCACATTGCTTTCATGAGTGGAAATCCATGTTCTGGACGCGGATTTTGTG 483
QY 481 GCCTTACTCTGTATAGATAAACAAGATCTAATTAATGTTGGACACTGCATCAATCAAT 540
DB 484 GCCTTACTCTGTATAGATAAACAAGATCTAATTAATGTTGGACACTGCATCAATCAAT 543
QY 541 TTTTGGAACTGAGTTTCAGTAACTAGTACCTGCGGATTCCTGCTCTCTCTCTCTCTCTCT 600
DB 544 TTTTGGAACTGAGTTTCAGTAACTAGTACCTGCGGATTCCTGCTCTCTCTCTCTCTCTCT 603
QY 601 ATATCTGGAACAAATCTCTCATGGATATAGAGATTCCTCGCATTTGTGCATGGCTGGTGTG 660
DB 604 ATATCTGGAACAAATCTCTCATGGATATAGAGATTCCTCGCATTTGTGCATGGCTGGTGTG 663
QY 661 CATCGAGAGTGTGTCTCAACACAGCTTTGGG 689
DB 664 CATCGAGAGTGTGTCTCAACACAGCTTTGGG 692
```

RESULT 4

```
QY 397 ATGAACCATTCATTAATCAAAAGCAATGAATACATTCATGCTGTTCTGAGTGAATC 456
DB 361 ATGAACCATTCATTAATCAAAAGCAATGAATACATTCATGCTGTTCTGAGTGAATC 420
QY 457 CATGTTTCTGGAGCGGAGATTTTGGGCTCATCTCTGTTTATAGATAAACAAGATCAAT 516
DB 421 CATGTTTCTGGAGCGGAGATTTTGGGCTCATCTCTGTTTATAGATAAACAAGATCAAT 480
QY 517 ACTTGTCTGGACACTGCATCCAAATTTTGGAACTGAGTTCAGTAACTAGTACCTCCAGCT 576
DB 481 ACTTGTCTGGACACTGCATCCAAATTTTGGAACTGAGTTCAGTAACTAGTACCTCCAGCT 540
QY 577 GGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 636
DB 541 GGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
QY 637 TCGCCATTCGTCATGCTGGTGTGTCATGAGGAGTGTCTTAAAGATTCCTTGGCTAACAAGCTC 756
DB 601 TCGCCATTCGTCATGCTGGTGTGTCATGAGGAGTGTCTTAAAGATTCCTTGGCTAACAAGCTC 720
QY 697 ATCAGTGTGTAAATAGTAAAGTATTCCTCTATATGAAAGTTCCTTGGCTAACAAGCTC 756
DB 661 ATCAGGTTGTAAATAGTAAAGTATTCCTCTATATGAAAGTTCCTTGGCTAACAAGCTC 720
QY 757 ACATCTGT--GGTGGGACACTTATCTACAAGTCTTTTACATTTAAGACAAAGTGGATG--TT 814
DB 721 CCATCTGTGGGGGACACTTATCTACAAGCTTTTACATTTAGAAAAGTGGATGCTT 780
QY 815 ATGGAACACTGGGGATGG---AGTCTGGTGTGATCGCGGATCCTCAAAATGAAGCAATCA 870
DB 781 ATGGAACACTGGGGATGGAAATCTGTGTGGATCGCGGAATCCTAAAATGAAGCAATCA 840
QY 871 TCTGTGTC--TGGAGTGGATGACACACACAGG--CAAGACAACAGTTGGAAA---CCCAA 923
DB 841 TCTGGGCTGTGAAGGGAGTGCACACCGGGGCAAGAACCGTTGGGAACCCCAAA 900
QY 924 AAAAGCCAGGCTGAAAAAAC 944
DB 901 AAAGGCCCGCTGAAAAAAC 921
```

RESULT 3
BF732384
LOCUS
DEFINITION
693 bp mRNA linear EST 30-MAR-2001
nae11a12.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3434806 3'
similar to TR:Q14089 Q14089 HYPOTHETICAL 40.0 KD PROTEIN ;, mRNA
sequence.
BF732384
VERSION
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 693)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christine Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LENL, send email to:
info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 490.
Location/Qualifiers

FEATURES

AI378788/c
LOCUS 712 bp mRNA linear EST 18-MAR-1999
DEFINITION tc67b04.x1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:2069647
3' similar to TR:Q14089 Q14089 HYPOTHETICAL PROTEIN ;, mRNA
sequence.
ACCESSION AI378788
VERSION AI378788.1 GI:4188641
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 712)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 861 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 461.
FEATURES
Location/Qualifiers
1..712
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2069647"
/clone_lib="Soares_NHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH108"
/note="Organ: mixed (see below); Vector: pTT73D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NBHPu, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following RAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT 166 a 143 c 173 g 230 t
ORIGIN
Query Match 30.0%; Score 684.8; DB 9; Length 712;
Best Local Similarity 97.6%; Pred. No. 1.4e-173;
Matches 695; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 889 GACCACAGGGCAAGACAGACAGTGGAAACCCAAAGCCAGGCTGAAAAACCTGGA 948
||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 712 GACCACCAAGCCAAAGAGACAGTGGAAACCCAAAGCCAGGCTGAAAAACCTGAC 653
QY 949 CCCTTTGGGCTGCTTTGGCACTGATGAATACCACTGGTTACAAATAGATTGTAATAG 1008
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 652 CGCTTTGGGCTGCTTTGGCACTGATGAATACCACTGGTTACAAATAGATTGTAATAG 593
QY 1009 GAAAAAATAATACAGGCATTATAACCACTGGATCCACCATGGTGGAGCACAATTACTAT 1068
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 592 GAAAAAATAATACAGGCATTATAACCACTGGATCCACCATGGTGGAGCACAATTACTAT 533
QY 1069 GTGTCCTGTACAGATCTGTACATGATGATGGCAGAAATGGACTGTGTACAGAGAG 1128
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 532 GTGTCCTGTACAGATCTGTACATGATGATGGCAGAAATGGACTGTGTACAGAGAG 473
QY 1129 CTTGGTGTGGAGCAAGATAGATTTCAGGAAACAAAGATTATCACAGGATGGCT 1188
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 472 CTTGGTGTGGAGCAAGATAGATTTCAGGAAACAAAGATTATCACAGGATGGCT 413
QY 1189 AATACTTTTGGCACCATTATTTGACGCTTTATTAGAGTGAATCTACCCAATGGCAG 1248
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 412 AATAACTTTTGGCACCATTATTGTCACGTTTTATTAGAGTGAATCCTACCCAATGGCAG 353
QY 1249 CAGAAATTTGCCATGAAATGAGCTGCTCGGATGTCAGTTTATTCTTAAAGTCTCTCT 1308
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 352 CAGAAATTTGCCATGAAATGAGCTGCTCGGATGTCAGTTTATTCTTAAAGTCTCTCT 293
QY 1309 CCAAACTTACTCAACCTCCAGCTCTCGGACAGCAATGACCTCAAAACACTACAGCC 1368
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 292 CCAAACTTACTCAACCTCCAGCTCTCGGACAGCAATGACCTCAAAACACTACAGCC 233
QY 1369 CTTCAAAATAGCCAAAGGTGTCGCCCAAAATTTACGCAACCTACAACTCGAGT 1428
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 232 CTTCAAAATAGCCAAAGGTGTCGCCCAAAATTTACGCAACCTACAACTCGAGT 173
QY 1429 ASCAATGAATTTCTTCACAGACAGCAACAACTGCCAGTCTCTGATATCAGAAATACT 1488
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 AGCAATGAATTTCTTCACAGACAGCAACAACTGCCAGTCTCTGATATCAGAAATACT 113
QY 1489 ACCGTAACCTCCAAATGAACCAAGATGAGCGCTGGCTGAGTCTTGTCCCTGTGCTG 1548
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 112 ACCGTAACCTCCAAATGAACCAAGATGAGCGCTGGCTGAGTCTTGTCCCTGTGCTG 53
QY 1549 GTCATGGTCTCCTCACTACTCTCATTTTCATATTAGTGTGCTTGGCACTGGA 1600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 52 GTCATGGTCTCCTCACTACTCTCATTTTCATATTAGTGTGCTTGGCACTGGA 1
RESULT 5
BI860608 603386804F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395609 5',
LOCUS mRNA sequence.
DEFINITION
ACCESSION BI860608
VERSION BI860608.1 GI:16001355
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 694)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12008 row: g column: 02
High quality sequence stop: 692.
FEATURES
Location/Qualifiers
1..694
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5395609"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: Sall; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT 230 a 165 c 147 g 152 t
ORIGIN
Query Match 30.0%; Score 683.4; DB 13; Length 694;
Best Local Similarity 99.1%; Pred. No. 3.4e-173;
Matches 687; Conservative 0; Mismatches 6; Indels 0; Gaps 0;


```

QY 419 AAGGCAATGAATACATCTGTTTCATGAGTGAATCCATGTTTCTGACCGGATTTT 478
Db 63 AAGGCAATGAATACATAGCTGTTTCATGAGTGAATCCATGTTTCTGACCGGATTTT 122
QY 479 TGGCCTCATACCTCTGTATAGATAAACAAGATCTAATTAATGTTTGGACATGATCCA 538
Db 123 TGGCCTCATACCTCTGTATAGATAAACAAGATCTAATTAATGTTTGGACATGATCCA 182
QY 539 ATTTTGTGAACCTGAGTTACAGTACTGCCAGCTGGTGTCTGCTTCTTCTTTGCTG 598
Db 183 ATTTTGTGAACCTGAGTTACAGTACTGCCAGCTGGTGTCTGCTTCTTCTTTGCTG 242
QY 599 AGATATCTGGACAATTCCTCATGATATAGAGATTCCTCGCAATGTCATGCTGCTG 658
Db 243 AGATATCTGGACAATTCCTCATGATATAGAGATTCCTCGCAATGTCATGCTGCTG 302
QY 659 TGCATGCAGAGTAGTGTCAACACAGTTGGCGGGCCCAATCAGTGTGTAATTAGTAAG 718
Db 303 TGCATGCAGAGTAGTGTCAACACAGTTGGCGGGCCCAATCAGTGTGTAATTAGTAAG 362
QY 719 GATTCCTCTATTATGAAGTCTCTTGGCTAACAAGCTCACATCTGCTGGTGACACTTAT 778
Db 363 GATTCCTCTATTATGAAGTCTCTTGGCTAACAAGCTCACATCTGCTGGTGACACTTAT 422
QY 779 CTACAAGCTCTTTTACATTTAACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTG 838
Db 423 CTACAAGCTCTTTTACATTTAACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTG 482
QY 839 GTGTGATCGGATCTCTCAAAATACAGCATCATCTGCTGAGTGGAGTGCACACACAG 898
Db 483 GTGTGATCGGATCTCTCAAAATACAGCATCATCTGCTGAGTGGAGTGCACACACAG 542
QY 899 GGCAAGAGAACAGTTGGAAACCCCAAAAGCCAGGCTGAAACAACTGGACCGCTTGGG 958
Db 543 GGCAAGAGAACAGTTGGAAACCCCAAAAGCCAGGCTGAAACAACTGGACCGCTTGGG 602
QY 959 CTGCTTTTGCCACTGATGAATACCACTGTTTACAATAGATTTGAATAAGGAAA 1012
Db 603 CTGCTTTTGCCACTGATGAATACCACTGTTTCCAATAGATTTGATTACGGAAA 656

RESULT 13
AA147037/c 671 bp mRNA linear EST 05-DEC-1996
LOCUS z032a05.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
DEFINITION IMAGE:588560 3' similar to TR:G704441 G704441 HYPOTHETICAL 40.0 KD
PROTEIN ; mRNA sequence.
ACCESSION AA147037.1 GI:1716444
VERSION AA147037.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 671)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapella,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisan,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham

```

```

FEATURES
source
High quality sequence stop: 439.
Location/Qualifiers
1..671
/organism="Homo sapiens"
/db_xref="GDB:4620862"
/db_xref="taxon:9606"
/clone="IMAGE:588560"
/clone_lib="Stratagene colon (#937204)"
/tissue_type="tumor"
/cell_line="T84 carcinoma cell line"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: colon; Vector: pBluescript SK-; Site:1: EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo dt. T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'"
BASE COUNT 176 a 162 c 133 g 198 t 2 others
ORIGIN
Query Match 25.6%; Score 584.8; DB 9; Length 671;
Best Local Similarity 97.9%; Pred. No. 1.4e-146;
Matches 645; Conservative 0; Mismatches 8; Indels 6; Gaps 5;
QY 590 CTTTTCGTGAGATATCTCGAACAATTCCTCATGATATAGATTCCTCGCCATTTGCA 649
Db 653 CTTTTCGTGAGATATCTCG-ACCATTCTCATGGATAT-GAGATTCCTCGCCAT--GTCTN 598
QY 650 TGGCTGGTGTGCATCGAGGAGTAGTCAACACAGTTGGCGGCCCAAAATCAGTGTGTAA 709
Db 597 TGGCTGGTGT-CATCGAGGAGTAGTGTCAACACAGTT-GCGCGCCAAATCAGTGTGTAA 540
QY 710 TTAGTAAAGGATTCCTCTATTATGAAGTCTTTTGGCTAACACGTCACATCTGTGTGG 769
Db 539 TTAGTAAAGGATTCCTCTATTATGAAGTCTTTTGGCTAACACGTCACATCTGTGTGG 480
QY 770 GACATCTATCTACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACTGGGA 829
Db 479 GACATCTATCTACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACTGGGA 420
QY 830 TGGAGTCTGGTGTGATCGCGGATCCTCAATACACAGCATCATCTGTGTGGAGTGGACTG 889
Db 419 TGGAGTCTGGTGTGATCGCGGATCCTCAATACACAGCATCATCTGTGTGGAGTGGACTG 360
QY 890 ACCACACAGGGCAAGAGAACAGTGTGAAACCCCAAAAGCCAGGCTGAAAAACCTGGAC 949
Db 359 ACCACACAGGGCAAGAGAACAGTGTGAAACCCCAAAAGCCAGGCTGAAAAACCTGGAC 300
QY 950 CGCCTTGGGCTGCTTTTGCCACTGATGAATACCACTGTTTACAAATAGATTTGAATAAG 1009
Db 299 CGCCTTGGGCTGCTTTTGCCACTGATGAATACCACTGTTTACAAATAGATTTGAATAAG 240
QY 1010 AAAAGAAAATACAGGCATTTAAACCCTGGATCCACCTGTTGGAGGACACAAATTAATG 1069
Db 239 AAAAGAAAATACAGGCATTTAAACCCTGGATCCACCTGTTGGAGGACACAAATTAATG 180
QY 1070 TCTCTGCCCTACAGATTCCTGTACAGTGTATGTTGGCAGAAATGCACTGTGTACAGAGAGC 1129
Db 179 TGTCTGCCCTACAGATTCCTGTACAGTGTATGTTGGCAGAAATGCACTGTGTACAGAGAGC 120
QY 1130 CTGGTGTGGAGCAAGATAAGATATTTCAGAGAAACAAGATTATCACAGGATGTGCGTA 1189
Db 119 CTGGTGTGGAGCAAGATAAGATATTTCAGAGAAACAAGATTATCACAGGATGTGCGTA 60
QY 1190 ATAACTTTTGGCCACCAATTTATGACGTTTATTATGATGATGATCTACCAATGCGAG 1248
Db 59 ATAACTTTTGGCCACCAATTTATGACGTTTATTATGATGATGATCTACCAATGCGAG 1
RESULT 14
BE894226 618 bp mRNA linear EST 20-OCT-2000
LOCUS BE894226
DEFINITION 601437167Fl NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922169 5',

```

```

mRNA sequence.
ACCESSION BE894226
VERSION BE894226.1 GI:10356380
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 618)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue procurement: ATCC/DCTD/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM9756 row: e column: 18
High quality sequence stop: 618.
Location/Qualifiers
1. 618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3922169"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="PH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 201 a 135 c 139 g 143 t
ORIGIN
Query Match 25.6%; Score 584; DB 12; Length 618;
Best Local Similarity 99.7%; Pred. No. 2.3e-146;
Matches 606; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
-QY 782 CAGTCTTTTACATTTAAGACAAAGTGGATGTTATGGAACACTGGGAGTGAGTCTGGTG 841
Db 1 CAAAGTCTTTTACATTTAAGACAAAGTGGATGTTATGGAACACTGGGAGTGAGTCTGGTG 60
QY 842 TGATCGGGATCCTCAATAACACCATCATCTGCTGCGAGTGCAGTACACACACAGGC 901
Db 61 TGATCGGGATCCTCAATAACACCATCATCTGCTGCGAGTGCAGTACACACACAGGC 120
QY 902 AAGAGAACAGTGGAAACCCAAAAGCCAGGCTGAAAAACCTGGACCGCTTGGGCTG 961
Db 121 AAGAGAACAGTGGAAACCCAAAAGCCAGGCTGAAAAACCTGGACCGCTTGGGCTG 180
QY 962 CTTTGGCACTGATGAATACACAGTGGTTACAATAGATTTGAATAAGAAAGAAATAA 1021
Db 181 CTTTGGCACTGATGAATACACAGTGGTTACAATAGATTTGAATAAGAAAGAAATAA 240
QY 1022 CAGGCATTATACCACTGGATCCACCAGTGGTGGAGCACATTAATCTAGTCTGCCATACA 1081
Db 241 CAGGCATTATACCACTGGATCCACCAGTGGTGGAGCACATTAATCTAGTCTGCCATACA 300
QY 1082 GAATCCTGTACAGTGATGGGAGAAATGGACTGTGTACAGAGAGCGTGTGGTGGAGC 1141
Db 301 GAATCCTGTACAGTGATGGGAGAAATGGACTGTGTACAGAGAGCGTGTGGTGGAGC 360
QY 1142 AAGATAAGATATTTCAAGGAACAAAGATATTATCACAGGATGTGCGTAATAACTTTTGC 1201
Db 361 AAGATAAGATATTTCAAGGAACAAAGATATTATCACAGGATGTGCGTAATAACTTTTGC 420
QY 1202 CA-CCAATTATGACGCTTTTATTAGAGTGAATCCTACCCCAATGGCAGCAGAAATGCC 1260
Db 421 CACCCCAATTATGACGCTTTTATTAGAGTGAATCCTACCCCAATGGCAGCAGAAATGCC 480

```

mRNA sequence.
 BE894226
 BE894226.1 GI:10356380
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 618)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue procurement: ATCC/DCTD/DTP
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLM9756 row: e column: 18
 High quality sequence stop: 618.
 Location/Qualifiers
 1. 618
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3922169"
 /clone_lib="NIH_MGC_72"
 /tissue_type="melanotic melanoma"
 /lab_host="PH10B (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site.1: NotI;
 Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

BASE COUNT 201 a 135 c 139 g 143 t
 ORIGIN
 Query Match 25.6%; Score 584; DB 12; Length 618;
 Best Local Similarity 99.7%; Pred. No. 2.3e-146;
 Matches 606; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

-QY 782 CAGTCTTTTACATTTAAGACAAAGTGGATGTTATGGAACACTGGGAGTGAGTCTGGTG 841
 Db 1 CAAAGTCTTTTACATTTAAGACAAAGTGGATGTTATGGAACACTGGGAGTGAGTCTGGTG 60
 QY 842 TGATCGGGATCCTCAATAACACCATCATCTGCTGCGAGTGCAGTACACACACAGGC 901
 Db 61 TGATCGGGATCCTCAATAACACCATCATCTGCTGCGAGTGCAGTACACACACAGGC 120
 QY 902 AAGAGAACAGTGGAAACCCAAAAGCCAGGCTGAAAAACCTGGACCGCTTGGGCTG 961
 Db 121 AAGAGAACAGTGGAAACCCAAAAGCCAGGCTGAAAAACCTGGACCGCTTGGGCTG 180
 QY 962 CTTTGGCACTGATGAATACACAGTGGTTACAATAGATTTGAATAAGAAAGAAATAA 1021
 Db 181 CTTTGGCACTGATGAATACACAGTGGTTACAATAGATTTGAATAAGAAAGAAATAA 240
 QY 1022 CAGGCATTATACCACTGGATCCACCAGTGGTGGAGCACATTAATCTAGTCTGCCATACA 1081
 Db 241 CAGGCATTATACCACTGGATCCACCAGTGGTGGAGCACATTAATCTAGTCTGCCATACA 300
 QY 1082 GAATCCTGTACAGTGATGGGAGAAATGGACTGTGTACAGAGAGCGTGTGGTGGAGC 1141
 Db 301 GAATCCTGTACAGTGATGGGAGAAATGGACTGTGTACAGAGAGCGTGTGGTGGAGC 360
 QY 1142 AAGATAAGATATTTCAAGGAACAAAGATATTATCACAGGATGTGCGTAATAACTTTTGC 1201
 Db 361 AAGATAAGATATTTCAAGGAACAAAGATATTATCACAGGATGTGCGTAATAACTTTTGC 420
 QY 1202 CA-CCAATTATGACGCTTTTATTAGAGTGAATCCTACCCCAATGGCAGCAGAAATGCC 1260
 Db 421 CACCCCAATTATGACGCTTTTATTAGAGTGAATCCTACCCCAATGGCAGCAGAAATGCC 480

QY 1261 ATGAAATGAGCTGCTCGATGTCAGTTTATTCCTAAAGGTGCTCTCCAAACTTACT 1320
 Db 481 ATGAAATGAGCTGCTCGATGTCAGTTTATTCCTAAAGGTGCTCTCCAAACTTACT 540
 QY 1321 CACCTCCACCTCCTCGGAACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAATA 1380
 Db 541 CAA-CTCCACCTCCTCGGAACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAATA 599
 QY 1381 GCCAAGG 1388
 Db 600 GCCAAGG 607
 RESULT 15
 AK006805 1272 bp mRNA linear HTC 19-JAN-2002
 Mus musculus adult male testis cDNA, RIKEN full-length enriched
 library, clone:1700055P21:endothelial and smooth muscle
 cell-derived neuropilin-like molecule, full insert sequence.
 AK006805
 AK006805.1 GI:12840068
 HTC; CAP trapper.
 Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
 clone_lib:RIKEN full-length enriched mouse cDNA library
 clone:1700055P21.
 Mus musculus
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 JOURNAL 99279253
 MEDLINE 10349636
 PUBMED
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 Prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 JOURNAL 20499374
 MEDLINE 11042159
 PUBMED
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multiplexed capillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL 20530913
 MEDLINE 11076861
 PUBMED
 4
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleischer, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
 Quackenbush, J., Schriml, L., Stauble, F., Suzuki, R., Tomita, M.,
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Anoh, H.,
 Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
 Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
 Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
 Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P.,
 Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:39:10 ; Search time 25.0427 Seconds
(without alignments)
3474.573 Million cell updates/sec

Title: US-10-060-830-1114

Perfect score: 3470

Sequence: 1 MPEFLLLVLLVLLLEDAGA.....TOEVSAGRGDGECDVKEIL 653

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3369	97.1	669	21 AAY70539	Human Factor 8 Hom
2	2340	67.4	583	22 AAG75450	Human colon cancer
3	1203.5	34.7	385	23 AAE22716	Human Neuropilin-H
4	1203.5	34.7	385	23 AAB79460	Human Neuropilin-H
5	1203.5	34.7	385	23 AAB79460	Novel human protei
6	1150.5	33.2	365	22 AAE22721	Human Neuropilin-H
7	1039	29.9	197	22 AAG01298	Novel human diagno
8	873.5	25.2	715	22 AAU00670	Human TANGO 229 po
9	756	21.8	487	22 AAU00628	Novel human protei
10	756	21.8	539	22 AAU00630	Novel human protei

11	756	21.8	586	22 AAU00629	Novel human protei
12	416	12.0	503	21 AAB19126	Polyptide isolat
13	398	10.9	398	23 AAE22715	Human Neuropilin-H
14	378	10.9	398	23 AAB79459	Human Neuropilin-H
15	355.5	10.2	2183	21 AAB03533	Murine factor V SE
16	355	10.2	480	23 AAU98718	Human integrin bin
17	352	10.1	480	18 AAW10364	Mouse developmenta
18	352	10.1	480	20 AAW94683	Murine Del-1 prote
19	352	10.1	513	20 AAW94685	Human Del-1 protei
20	351	10.1	2224	20 AAY49564	Human lipoprotein
21	349	10.1	159	22 AAB71907	Human FV C2 domain
22	347	10.0	2224	17 AAW04254	Human Factor V. H
23	346	10.0	481	18 AAW10365	Human developmenta
24	343	9.9	217	16 AAR77253	HMFG 46 kDa antige
25	343	9.9	218	16 AAR77254	HMFG 46 kDa antige
26	343	9.9	387	16 AAR77252	HMFG 46 kDa antige
27	343	9.9	387	21 AAY94453	Human lactadherin
28	330	9.5	321	20 AAW94697	Human milk fat glo
29	323.5	9.3	426	21 AAY94454	Mouse lactadherin
30	317	9.1	1431	22 AAB67960	Amino acid sequenc
31	317	9.1	2343	20 AAW80989	Canine factor VIII
32	317	9.1	2343	21 AAY57846	Canine factor VIII
33	313.5	9.0	160	22 AAB71905	Canine FVIII C2 do
34	313	9.0	320	20 AAW98891	Peptide SEQ ID NO:
35	312.5	9.0	2332	23 AAU79874	Human factor VIII
36	309.5	8.9	2332	23 AAU79875	Human factor VIII
37	309.5	8.9	2351	8 AAF70448	Human factor VIII:
38	309.5	8.9	2351	16 AAR78223	Human Factor-VIII:
39	308.5	8.9	684	16 AAR73022	Human Factor-VIII
40	308.5	8.9	684	16 AAR74091	Factor-VIII light
41	308.5	8.9	1014	8 AAF71139	Factor-VIII:c varia
42	308.5	8.9	1383	18 AAW33227	Procoagulant-activ
43	308.5	8.9	1383	18 AAW33228	Procoagulant-activ
44	308.5	8.9	1383	18 AAW33229	Procoagulant-activ
45	308.5	8.9	1424	9 AAP80268	Modified factor VI

ALIGNMENTS

RESULT 1
AAY70539
ID AAY70539 standard; Protein; 669 AA.

XX AAY70539;
AC
XX
DT 04-JUL-2000 (first entry)
XX
DE Human Factor 8 Homologue.
XX
Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW cerebroprotective; therapeutic; coagulation related disorder;
KW haemophilia; stroke; screening.
XX
XX Homo sapiens.
OS
XX
XX WO200012532-A1.
XX
PD 09-MAR-2000.
XX
PF 20-AUG-1999; 99WO-US19047.
XX
PR 31-AUG-1998; 98US-0098521.
XX
XX (ELIL) LILLY & CO ELI.

XX Rostock PRJ, Su W, Li XM;
XX WPI: 2000-256580/22.
XX N-PSDB; AA251872.
XX
XX Factor 8 homolog polypeptides and nucleic acids encoding them for
XX treating coagulation related disorders such as hemophilia and stroke
PT

XX Claim 3; Page 64-66; 68pp; English.
XX The present sequence is a human Factor 8 homologue (F8H),
CC a coagulation cofactor which is selectively expressed in
CC haematopoietic, heart and reproductive tissues. It has haemostatic and
CC cerebroprotective activities. The F8H contains a Factor 5/8 signature
CC and is useful as a therapeutic for treating coagulation related diseases
CC such as haemophilia and stroke. The nucleic acid is useful as
CC hybridisation probe and amplification primer for detecting deficiencies
CC in the level of F8H mRNA, for screening F8H gene mutations and for
CC monitoring regulation of gene expression. Fragments of the nucleic acid
CC are also useful as diagnostic probes and primers, and can be used in
CC screening methods such as those using DNA chips. The present sequence is
CC also useful as a target to screen therapeutically useful modulators
CC of the F8H.

XX Sequence 669 AA;

Query Match 97.1%; Score 3369; DB 21; Length 669;
Best Local Similarity 100.0%; Pred. No. 8e-282;
Matches 631; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GYCGGLGLOMNSIESKSGNEITLLFMGSHVSGRFLASYSVIDKODLITCLDTASNFLE 82
DB 39 GYCGGLGLOMNSIESKSGNEITLLFMGSHVSGRFLASYSVIDKODLITCLDTASNFLE 98
QY 83 PEFKSKYCPAGCLLPFAEISGTIPHGYSRSPCLCMAGVHAGVVSNTLGGQISVVISKIPY 142
DB 99 PEFKSKYCPAGCLLPFAEISGTIPHGYSRSPCLCMAGVHAGVVSNTLGGQISVVISKIPY 158
QY 143 YESSLANNVTSVVGHLSTSLFTFKTSGCYGTGMSGVADPQITASSVLEWTDHTGQEN 202
DB 159 YESSLANNVTSVVGHLSTSLFTFKTSGCYGTGMSGVADPQITASSVLEWTDHTGQEN 218
QY 203 SWPKPKARLKPKPPWAAATDEYQWLQDLNKRKITGIITGSTMVHNHYVSAYRIL 262
DB 219 SWPKPKARLKPKPPWAAATDEYQWLQDLNKRKITGIITGSTMVHNHYVSAYRIL 278
QY 263 YSDDGKWTYRPPGVQEKIFQGNKDYHQDVNRNPLPPIIARFIRVNPQWQKIAMKM 322
DB 279 YSDDGKWTYRPPGVQEKIFQGNKDYHQDVNRNPLPPIIARFIRVNPQWQKIAMKM 338
QY 323 ELGCGQIFPKGRPKLTQPPPPNSDLKNTTAPPKIAGRAPKFTQPLQPRSSNEFFPAQ 382
DB 339 ELGCGQIFPKGRPKLTQPPPPNSDLKNTTAPPKIAGRAPKFTQPLQPRSSNEFFPAQ 398
QY 383 TEQTASPDIRNTVTPNVTKDVALAALVPLVMVLTLLILVCAWHNRKKKTEGT 442
DB 399 TEQTASPDIRNTVTPNVTKDVALAALVPLVMVLTLLILVCAWHNRKKKTEGT 458
QY 443 YDLPYWDRAWGMKQFLPAKAVDHEETPVRYSSSEVNHLSPREVTTLQADSAEYAPQ 502
DB 459 YDLPYWDRAWGMKQFLPAKAVDHEETPVRYSSSEVNHLSPREVTTLQADSAEYAPQ 518
QY 503 LVGCGVGLTHQRTFPEEGKEAGYADLDYNSPGQEVHAYAEPLPITGPEYATPIIMD 562
DB 519 LVGCGVGLTHQRTFPEEGKEAGYADLDYNSPGQEVHAYAEPLPITGPEYATPIIMD 578
QY 563 MSCHPTTSVQGPSTSTFKATGNOPPLVGTNTLLSRTDSCSSAAQAYDTPKAGKPLPA 622
DB 579 MSCHPTTSVQGPSTSTFKATGNOPPLVGTNTLLSRTDSCSSAAQAYDTPKAGKPLPA 638
QY 623 PDELVTYQVQPOSTQEVSGAGRGDCDFKEIL 653
DB 639 PDELVTYQVQPOSTQEVSGAGRGDCDFKEIL 669

RESULT 2
AAG75450
ID AAG75450 standard; Protein; 583 AA.
XX
AC AAG75450;

XX 03-SEP-2001 (first entry)
XX Human colon cancer antigen protein SEQ ID NO:6214.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26524.
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
XX N-PSDB; AAH34855.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 7657-7660; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAB77789 represent sequences used in the exemplification of the
XX present invention.
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 583 AA;
Query Match 67.4%; Score 2340; DB 22; Length 583;
Best Local Similarity 84.6%; Pred. No. 4.7e-193;
Matches 451; Conservative 2; Mismatches 4; Indels 76; Gaps 1;
QY 1 MPFLFLLLVLLVLLLEDAGAQQ----- 22
DB 25 MPFLFLLLVLLVLLLEDAGAQQDGGCHTVLGPESGTLTSINYPQTYPNSTVCWEIRVK 84
QY 23 -----GKYCGIGLOMNSIESKSGNEIT 44
DB 85 MGERVRIKFGDFIEDSDSCHFNLYRIYNGIVSRTEIGKYCGIGLOMNSIESKSGNEIT 144
QY 45 LLFMSGIHVSGRFLASYSVIDKODLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 104
DB 145 LLFMSGIHVSGRFLASYSVIDKODLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 204
QY 105 PHGYRDSPLCMAGVHAGVVSNTLGGQISVVISKIPYESSLANNVTSVVGHLSTSLFT 164
DB 205 PHGYRDSPLCMAGVHAGVVSNTLGGQISVVISKIPYESSLANNVTSVVGHLSTSLFT 264

QY 165 FKTSGCYTGLMGESGVIADPOITASSVLEWTDHTGQENSWPKPKARLKKPGPPWAAAFATD 224
DB 265 FKTSGCYTGLMGESGVIADPOITASSVLEWTDHTGQENSWPKPKARLKKPGPPWAAAFATD 324
QY 225 EYQWLQIDLNKKEKITGITGTSMTVEHNYVSAYRILYSDGOKWTYVREPGVEQDKIF 284
DB 325 EYQWLQIDLNKKEKITGITGTSMTVEHNYVSAYRILYSDGOKWTYVREPGVEQDKIF 384
QY 295 QGNKDYHQDVNRNLPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPPKLTQPPPP 344
DB 385 QGNKDYHQDVNRNLPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPPKLTQPPPP 444
QY 345 RNSNDLKNTPAPPKIAKRAPKFTQPLQPRSSNEFFPAQTEQTTASPDTRNTVTPNWKD 404
DB 445 RNSNDLKNTPAPPKIAKRAPKFTQPLQPRSSNEFFPAQTEQTTASPDTRNTVTPNWKD 504
QY 405 VALAAVLPVLVWLTLLILVCAWHNRNKKKTEGYDLPYWDRAQWKGK 457
DB 505 VALAAVLPVLVWLTLLILVCAWHNRNKKKTEGYDLPYWDRAQWKGK 557

RESULT 3

ID AAE22716 standard; Protein; 385 AA.
XX AAE22716;
AC AAE22716;
XX AAE22716;
DT 09-AUG-2002 (first entry)
XX Human neuropilin-Hy2 protein.
DE Human neuropilin-Hy2 protein.
XX Human neuropilin-Hy2 protein.

Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
thrombocytopenia; memory; platelet; plastic anaemia; antinflammatory;
ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
wound healing; tissue repair; Parkinson's disease; Huntington's disease;
amyotrophic lateral sclerosis; Shy-brager syndrome; trauma; spinal cord;
cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
systemic lupus erythematosus; rheumatoid arthritis; antichyroid; cancer;
autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
nootropic; neuroprotective; vulnary; anticonvulsant; antiparasitic;
cerebroprotective; tranquiliser; virucide; antibacterial; cytostatic;
immunosuppressive; chromosome 6q21.

Homo sapiens.

Key Location/Qualifiers
FT Peptide 1..20
FT /label= signal_peptide
FT Protein 61..385
FT /note= "Human mature neuropilin-Hy2 protein"

XX WO200222815-A1.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28488.

XX 11-SEP-2000; 2000US-0659671.

XX 06-SEP-2001; 2001US-317902P.

XX (HYSE-) HYSEQ INC.

XX Tang YF;

XX WPI; 2002-393966/42.

XX N-PSDB; AAD35994.

XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects

XX Claim 3; Page 128-130; 152pp; English.

The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
like polypeptides and polynucleotides are useful in modulating neuronal
growth regenerative capacity, treating neurodegenerative diseases,
diagnosing and mapping genetic neuronal defects and degenerative diseases
like Alzheimer's disease and for treating learning and memory disorders.
They are also useful for inducing angiogenesis, neovascularisation, as
well as organ growth and development e.g. heart and other tissues.
Antagonists of neuropilin-like polypeptides are useful for treating
cancers and other malignant diseases. Neuropilin is used to treat
platelet disorders e.g. thrombocytopenia, plastic anaemia and paroxysmal
nocturnal haemoglobinuria and is used in nerve tissue growth or
regeneration, in wound healing, tissue repair and replacement and in
healing of bones, incisions and ulcers. Compositions comprising the
sequences of the invention are useful for treating diseases of peripheral
nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
stroke, ulcers, immune deficiencies and immune disorders, infections by
hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
inflammatory eye disease. The nucleic acids of the invention are used in
gene therapy techniques. The present sequence is human neuropilin-Hy2
protein. Neuropilin-Hy2 gene is located on chromosome 6q21.

XX Sequence 385 AA;

Query Match 34.7%; Score 1203.5; DB 23; Length 385;
Best Local Similarity 65.3%; Pred. No. 3.5e-95;
Matches 254; Conservative 11; Mismatches 37; Indels 87; Gaps 4;

QY 1 MPFLFLLLVLLLEDDAGAAQ-----GKVGGLQNMHSIESKGNIT 44
DB 1 MPFLFLLLVLLLEDDAGAAQGGCGHTVLPGESGTLTSINYPQYPSNVTCEWIRVK 60
QY 23 -----GKVGGLQNMHSIESKGNIT 44
DB 61 MGERVRIKFGDFIEDSDSCHFNLYRIYNGIVSRTEIGKYCGLGLQNMHSIESKGNIT 120
QY 45 LLEFMSGIIHVSGRGLASVVDKQDLITCLDTASNFLEPFSKYCPAGCLLPFAISGTI 104
DB 121 LLEFMSGIIHVSGRGLASVVDKQDLITCLDTASNFLEPFSKYCPAGCLLPFAISGTI 180
QY 105 PHGYRDSPLCMAGVHAGVVSNTLGGQISVWISKIPYESSLANNVTSVVGHLSTSLFT 164
DB 181 PHGYRDSPLCMAGVHAGVVSNTLGGQISVWISKIPYESSLANNVTSVVGHLSTSLFT 240
QY 165 FKTSGCYTGLMGESGVIADPOITASSVLEWTDHTGQENSWPKPKARLKKPGPPWAAAFATD 224
DB 241 FKTSGCYTGLMGESGDRGSSNNSTVLEWTDHTGQENSWPKPKARLKKPGPPWAAAFATD 300
QY 225 EYQWLQIDLNKKEKITGITGTSMTVEHNYVSAYRILYSDGOKWTYVREPGVE 279
DB 301 EYQWLQIDLNKKEKITGITGTSMTVEHNYVSAYRILYSDGOKWTYVREPGVE 355
QY 280 QDKIFQGNKD-YHODVRNPNFLPPIIARFI 307
DB 356 QDKIFQGNKRIITRMVRNPNFLPPIIARLL 384

RESULT 4

ID AAU79460 standard; Protein; 385 AA.
XX AAU79460;
AC AAU79460;
XX AAU79460;
DT 15-JUL-2002 (first entry)


```
XX Claim 20; SEQ ID NO 654; 509pp; English.
XX
XX The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention.
XX
XX Sequence 385 AA;
XX
XX Query Match 34.7%; Score 1203.5; DB 23; Length 385;
XX Best Local Similarity 65.38; Pred. No. 3.5e-95;
XX Matches 254; Conservative 11; Mismatches 37; Indels 87; Gaps 4;
XX
XX 1 MFLFLLLLVLLLEDAQAQ-----GKYGGLQMNHSTESKGNIT 44
XX |||||
XX 1 MFLFLLLLVLLLEDAQAQCGGHTVLGPGESGTLTINYPQTPNSTVCWEIRVK 60
XX
XX 23 -----GKYGGLQMNHSTESKGNIT 44
XX |||||
XX 61 MGERVRIKFGDPDIEDSDSCHFNLYRIYNGIGVSRTEIGKYGGLQMNHSTESKGNIT 120
XX |||||
XX 45 LLFMSGIHVGSGFLASVIDKDLITCLDPTASNFLEPEFSKYCPAGCLLPFAEISGTI 104
XX |||||
XX 121 LLFMSGIHVGSGFLASVIDKDLITCLDPTASNFLEPEFSKYCPAGCLLPFAEISGTI 180
XX |||||
XX 105 PHGYRDSPLCMAGVHAGVSVNTLGGQISVLSKGIPEYESSLANNVTSVGHLSLFT 164
XX |||||
XX 181 PHGYRDSPLCMAGVHAGVSVNTLGGQISVLSKGIPEYESSLANNVTSVGHLSLFT 240
XX |||||
XX 165 FRTSGCYGTLMGSGVADPQITASSVLEWTDHTQGENSWKPKKRLKPGPPWAAAFATD 224
XX |||||
XX 241 FRTSGCYGTLMGSGDGRSSNITSVLEWTDHTQGENSWKPKKQAETWTALGAFATD 300
XX |||||
XX 225 FYQWLQIDLNKKKTKGIITGSTWVEHNYVSVARILYSDD-----GQKWTVYRPGVE 279
XX |||||
XX 301 FYQWLQIDLNKKKTKGIITGSTWVS-----TITMCLPTESCTVMGRNGLCTESLVVE 355
XX |||||
XX 280 QDKIFOGNKD-YHDOVRNLFPLPITAREI 307
XX |||||
XX 356 QDKIFOGNRIITRMVRNLFPLPITARLL 384
XX |||||
XX
XX RESULT 6
XX ID AAE22721
XX AC AAE22721 standard; Protein; 365 AA.
XX
XX DT 09-AUG-2002 (first entry)
XX
XX Human neuropilin-Hy2 mature protein.
XX
XX Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW opthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neurotropic; neuroprotective; vulnerary; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive.
```

```
XX OS Homo sapiens.
XX WO200222815-A1.
XX 21-MAR-2002.
XX
XX 12-SEP-2001; 2001WO-US28488.
XX
XX 11-SEP-2000; 2000US-0659671.
XX 06-SEP-2001; 2001US-317902P.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT;
XX
XX WPI; 2002-393966/42.
XX
XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
XX useful for treating neurodegenerative diseases e.g. Alzheimer's
XX disease, and for diagnosing and mapping genetic neuronal defects -
XX
XX Disclosure; Page 131-132; 152pp; English.
XX
XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
XX neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
XX like polypeptides and polynucleotides are useful in modulating neuronal
XX growth regenerative capacity, treating neurodegenerative diseases,
XX diagnosing and mapping genetic neuronal defects and degenerative diseases
XX like Alzheimer's disease and for treating learning and memory disorders.
XX They are also useful for inducing angiogenesis, neovascularisation, as
XX well as organ growth and development e.g. heart and other tissues.
XX Antagonists of neuropilin-like polypeptides are useful for treating
XX cancers and other malignant diseases. Neuropilin is used to treat
XX platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
XX nocturnal haemoglobinuria and is used in nerve tissue growth or
XX regeneration, in wound healing, tissue repair and replacement and in
XX healing of bones, incisions and ulcers. Compositions comprising the
XX sequences of the invention are useful for treating diseases of peripheral
XX nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
XX syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
XX spinal cord disorders, head trauma and cerebrovascular diseases e.g.
XX stroke, ulcers, immune deficiencies and immune disorders, infections by
XX hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
XX mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
XX multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
XX autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
XX mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
XX inflammatory eye disease. The nucleic acids of the invention are used in
XX gene therapy techniques. The present sequence is human neuropilin-Hy2
XX mature protein.
XX
XX Sequence 365 AA;
XX
XX Query Match 33.2%; Score 1150.5; DB 23; Length 365;
XX Best Local Similarity 79.7%; Pred. No. 1.2e-90;
XX Matches 232; Conservative 11; Mismatches 37; Indels 11; Gaps 3;
XX
XX 23 GKYCGGLQMNHSTESKGNITLLFMSTHVSGRFLASVIDKDLITCLDPTASNFLE 82
XX |||||
XX 79 GKYCGGLQMNHSTESKGNITLLFMSTHVSGRFLASVIDKDLITCLDPTASNFLE 138
XX |||||
XX 83 PFESKYCPAGCLLPFAEISGTIPHGVRDSSPLCMAGVHAGVSVNTLGGQISVVISKGPY 142
XX |||||
XX 139 PFESKYCPAGCLLPFAEISGTIPHGVRDSSPLCMAGVHAGVSVNTLGGQISVVISKGPY 198
XX |||||
XX 143 YESSLANNVTSVYVGHLSLFTFKTSKCYGTLMGSGVIADPQITASSVLEWTDHTGQEN 202
XX |||||
XX 199 YESSLANNVTSVYVGHLSLFTFKTSKCYGTLMGSGDGRSSNITSVLEWTDHTGQEN 258
XX |||||
XX 203 SWKPKKARLKKPGPPWAAAFATDEYQWLQIDLNKKKTKGIITGSTWVEHNYVSVARIL 262
XX |||||
XX 259 SWKPKKSAEKTWTALGAFATDEYQWLQIDLNKKKTKGIITGSTWVS-----TITMCL 313
XX |||||
```

QY 263 YSDD-----GQKWTYVREPQVQDKIFQGNKD-YHQDVYNNFLPPIIARFI 307
 Db 314 PTESCTVMGRNGLCTESLVVQDKIFQGNKRIITRMVYNNFLPPIIARLL 364

RESULT 7

ABG01298
 ID ABG01298 standard; Protein; 197 AA.

XX AC ABG01298;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #1289.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX XX WPI; 2001-639362/73.

DR N-PSDB; AAS65485.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID No 31657; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 197 AA;

Query Match 29.9%; Score 1039; DB 22; Length 197;
 Best Local Similarity 100.0%; Pred. No. 2.le-81;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 457 MKQFLPAKAVDHEETPVYSSSEVNHLSPREVTTVLQADSAEYAPQLVGGIVGTLHQST 516

Db 1 MKQFLPAKAVDHEETPVYSSSEVNHLSPREVTTVLQADSAEYAPQLVGGIVGTLHQST 60
 QY 517 FKPEEGKEAGYADLDPNYSPQGEVYHAYAEPLPITGPEYATPIIMDSHGHPITTSVGQPS 576
 Db 61 FKPEEGKEAGYADLDPNYSPQGEVYHAYAEPLPITGPEYATPIIMDSHGHPITTSVGQPS 120
 QY 577 STFKATGNQPPPLVGTNTLLSRDTSCSSAAQAYDTPKRAGKPLPAPDELVYQVQSTOE 636
 Db 121 STFKATGNQPPPLVGTNTLLSRDTSCSSAAQAYDTPKRAGKPLPAPDELVYQVQSTOE 180
 QY 637 VSGAGRDGECDFKEIL 653
 Db 181 VSGAGRDGECDFKEIL 197

RESULT 8

AAU00670

ID AAU00670 standard; Protein; 715 AA.

XX AC AAU00670;

XX DT 07-SEP-2001 (first entry)

XX DE Human TANGO 229 polypeptide.

XX KW Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
 KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
 KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
 KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease;
 KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
 KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
 KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
 KW muscular dystrophy; immuno-competence; vertebrate; blood; serum.

OS Homo sapiens.

XX PH Key Location/Qualifiers

FT Peptide 1..34

FT Domain /note= "Signal peptide"

FT Protein 35..455

FT Domain /note= "Extracellular domain"

FT Protein 35..715

FT Domain /note= "Mature human TANGO 229"

FT Domain 456..480

FT Domain /note= "Transmembrane domain"

FT Domain 481..715

FT Domain /note= "Cytoplasmic domain"

XX WO200129088-A1.

XX 26-APR-2001.

XX 23-JUN-2000; 2000WO-US17386.

XX 19-OCT-1999; 99US-0420707.

XX (MILL-) MILLENNIUM PHARM INC.

XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

XX WPI; 2001-308477/32.

DR N-PSDB; AAS00660.

XX New isolated nucleic acid molecule for diagnosis, prevention, and
 PT therapy of human and other animal disorder, or as modulating agent for
 PT regulating cellular processes -
 XX Claim 8; Fig 1; 263pp; English.
 XX The sequence represents human TANGO 229 polypeptide. This protein and
 CC similar others exhibit the ability to affect growth, proliferation,
 CC survival, differentiation, activity, morphology, or movement/migration

CC of, e.g. T cells and cells of the heart, liver, pancreas, placenta,
CC brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral
CC blood leukocyte, bone marrow or thymus tissue. They can be used as
CC modulating agents for regulating cellular processes, thus, the proteins
CC and their associated nucleic acids can be used to prognosticate, prevent,
CC diagnose, or treat disorders associated with physiological processes.
CC These disorders include abnormal blood coagulation, asthma, anaphylaxis,
CC hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria,
CC atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention
CC deficit disorder, Crohn's disease, gastroenteritis, goitre,
CC hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and
CC muscular dystrophy. Antibodies to disorders such as these can be made by
CC providing a polypeptide of the invention to an immuno-competent
CC vertebrate and harvesting blood or serum from the vertebrate.
XX
SQ Sequence 715 AA;
Query Match 25.2%; Score 873.5; DB 22; Length 715;
Best Local Similarity 35.2%; Pred. No. 2.8e-66;
Matches 225; Conservative 102; Mismatches 209; Indels 103; Gaps 22;
QY 13 LLELDAGAQKQKCYGLQMNHSIESKNEITLLEFMSGIHVSGRGLASYSVIDKQDLIT 72
DB 99 LLFTSSDQGPYCG-SMTVPKELLNTSEVTRFESGSHISGRGFLTYASSDHPDLIT 157
QY 73 CLDTASNLEPEFSKYCPAGCLLPFAEISGTIPHYRDSPLCMAGVHAGVNSNTLGGQI 132
DB 158 CLERASHYLKTEYSKFCPAGCRDVGADISGNMVDGYRDTSLCKAAIHAGIADDELGGQI 217
QY 133 SVVISKGIPIYESSLANNVTSVVGHLSTLFTFKTSGCYGTGLMESGVTDAPQITASSVL 192
DB 218 SVLQKGISRYEGILANGVLSRDGSLSKRFLFTSGCSRLSFE-----PDGOIRASSW 273
QY 193 EWTDTHTQENSKWKPKARLKKPGPPWAAATDE---YQWLQIDLNKKEKITGIITGST 248
DB 274 QSVNESGDQVHWSPPGARLODQGPSWASGDSNNHKKPREWLEIDLGEKKKITGIITGST 333
QY 249 MVEHNYYSAYRILYSDGQKVTYREPQVEODKIFQGNKDYHQDVNRNPLPPIARFIR 308
DB 334 QSNFNFYKSEFYNNFKNNNSKWKTYKGIYVNEEKVFQGSNFRDPQVNNFIPPIVARYVR 393
QY 309 VNPTQWQKIAMKMLGCOFTPKRPPKLTOTPPPRNSND---LKNTPAPPKIAKGRA 364
DB 394 VVPQTWHORIALUKVELIGQ-----ITO-----GNDSLVWRKTSQTSVSTKKED 438
QY 365 PRFTQLOPAPRSSNEPPAQTQETASPDIRNTVTPNVTKDVALAALVPLVNLVLTLL 424
DB 439 EITRPI-----PSEET--STGINITT-----VAIPLVLLVLFAGMGIFA 478
QY 425 ILVCAWHNRKKTGEGYDLPYW-----DRAGWKGMKQFLPAKAVDHEETP-VRYSSSE 479
DB 479 AF-----RKKKKGS---PYGSAEAKTDCWKQIKY-----PFARHQSAAE 515
QY 480 --VNHLSPREV---TVLQADSAEYQAQPLVGVGLHGRSTFKP--EGKEAGYADLDP 532
DB 516 FTISDNEKEMTKDLITSDMADTQOPLMIGTGTVRKGSFTRPMDTDAEAG-VSTDA 574
QY 533 ---YNSPGQEVYHAYAEPLITGPYATPIIMDSHGHTTSVQGPSTSTFKATGNQP--- 586
DB 575 GGHYDCPQAGRHEVALPLAPPEYATIV---ERHVLRAHTFSAQSGYRVPGPQGHK 631
QY 587 -----PPL--VGTYNLLSTRDSCSSAQAYDTPKA 615
DB 632 HSLSSGGFSPVAGQAQGDGYQRPHSAQPADRGDRPKA 670
RESULT 9
AAU00628
ID AAU00628 standard; Protein; 487 AA.
XX
AC AAU00628;
XX
DT 29-AUG-2001 (first entry)

XX Novel human protein (NHP) sequence #1.
DE
XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
XX NHP-mediated pathway.
OS Homo sapiens.
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 990US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX (LEXI-) LEXICON GENETICS INC.
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-290917/30.
DR N-PSDB; AAS00613.
XX
PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
PS Claim 4; Page 26; 33pp; English.
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
XX
SQ Sequence 487 AA;
Query Match 21.8%; Score 756; DB 22; Length 487;
Best Local Similarity 36.2%; Pred. No. 2.2e-56;
Matches 182; Conservative 87; Mismatches 156; Indels 78; Gaps 15;
QY 13 LLELDAGAQKQKCYGLQMNHSIESKNEITLLEFMSGIHVSGRGLASYSVIDKQDLIT 72
DB 47 LLFTSSDQGPYCG-SMTVPKELLNTSEVTRFESGSHISGRGFLTYASSDHPDLIT 105
QY 73 CLDTASNLEPEFSKYCPAGCLLPFAEISGTIPHYRDSPLCMAGVHAGVNSNTLGGQI 132
DB 106 CLERASHYLKTEYSKFCPAGCRDVGADISGNMVDGYRDTSLCKAAIHAGIADDELGGQI 165
QY 133 SVVISKGIPIYESSLANNVTSVVGHLSTLFTFKTSGCYGTGLMESGVTDAPQITASSVL 192
DB 166 SVLQKGISRYEGILANGVLSRDGSLSKRFLFTSGCSRLSFE---PDGOIRASSW 221
QY 193 EWTDTHTQENSKWKPKARLKKPGPPWAAATDE---YQWLQIDLNKKEKITGIITGST 248
DB 222 QSVNESGDQVHWSPPGARLODQGPSWASGDSNNHKKPREWLEIDLGEKKKITGIITGST 281
QY 249 MVEHNYYSAYRILYSDGQKVTYREPQVEODKIFQGNKDYHQDVNRNPLPPIARFIR 308
DB 282 QSNFNFYKSEFYNNFKNNNSKWKTYKGIYVNEEKVFQGSNFRDPQVNNFIPPIVARYVR 341

```
QY 309 VNPTQWQKIAMKMLLGCOFIPKGRPPKLTOTPPPRNSND-----LKNTPAPPKIAKGRA 364
Db 342 VVPOTWHORIALKVELIGCO-----ITQ-----GNDSLVWRKTSQSTSVSTKKED 386
QY 365 PKFTQLOPQRSSNEPPAQTOTASPDIRNTVTPNTKDVAAALVPLVLMVLTLLIL 424
Db 387 ETRITRPI-----PSEET--STGINITT-----VAIPLVLLVLFVAGMGIFA 426
QY 425 ILVCAWHNRKTKTEGYDLPYW-----DRAGWWMKQKFLPAKAVDHEETP-VRYSSSE 479
Db 427 AF-----RKKKKKGS---PYGSAAQKTCWKQIKY-----PFAHQSAE 463
QY 480 --VNHLSPREVTTLQADSAEA 500
Db 464 FTISYDNEKEMTKLDLITSDMA 486

RESULT 10
AAU00630
ID AAU00630 standard; Protein; 539 AA.
XX
AC AAU00630;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #3.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
XX
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX
DR N-PSDB; AAS00615.
XX
PT Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
PS Claim 5; Page 29-30; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
XX
```

```
SQ Sequence 539 AA:
Query Match 21.8%; Score 756; DB 22; Length 539;
Best Local Similarity 36.2%; Pred. No. 2.6e-56;
Matches 182; Conservative 87; Mismatches 156; Indels 78; Gaps 15;

QY 13 LLELDAGAQOQGYCGGLGLOMHSIESKGNELTLLFMSGHVSGRFLASVYDVKQDLIT 72
Db 99 LLFTSSSQOYGYCG-SMTVPKELLNTSEVTVRFESSHISGRGFLLLTYASSDHPDLIT 157
QY 73 CLDTASNFLEPFESKYPACGLLPFAEISGTIPHGVRDSSPLCMAGVHAGVSVNTLGGQI 132
Db 158 CLERASHYLKTEYSKFCPACGRDVGDISGNMVDGYRDTSLKAAIHAGIILADELGGQI 217
QY 133 SVVISKGIPIYESSLANNTVSVGHLSLSTLFTKTSCTGCTLGMESGVADPOITASSVL 192
Db 218 SVLQRKGISRYEGILANGVLSRDGSLSKRFLFTSNGCSRSLSE----PDGQIRASSW 273
QY 193 EWTDTGTGQENSMKPKKARLKPPWAAAFATDE---YQWLQIDLNKKKITGIITGST 248
Db 274 QSVNESGDQVHWSPGQARLDQGPSWASGDSSNNHKPREWLEIDLGEKKKITGIITGST 333
QY 249 MVEHNYYSAYRILYSDGQKWTYREPGVEQDKIPQGNKDYHQDVNRNFPPIIARFIR 308
Db 334 QSNFNFYKVSFVMNPKNNNSKNKTYKGIYNNEKVFQGNPRDPVQNNFIPPIVARYVR 393
QY 309 VNPTQWQKIAMKMLLGCOFIPKGRPPKLTOTPPPRNSND-----LKNTPAPPKIAKGRA 364
Db 394 VVPOTWHORIALKVELIGCO-----ITQ-----GNDSLVWRKTSQSTSVSTKKED 438
QY 365 PKFTQLOPQRSSNEPPAQTOTASPDIRNTVTPNTKDVAAALVPLVLMVLTLLIL 424
Db 439 ETRITRPI-----PSEET--STGINITT-----VAIPLVLLVLFVAGMGIFA 478
QY 425 ILVCAWHNRKTKTEGYDLPYW-----DRAGWWMKQKFLPAKAVDHEETP-VRYSSSE 479
Db 479 AF-----RKKKKKGS---PYGSAAQKTCWKQIKY-----PFAHQSAE 515
QY 480 --VNHLSPREVTTLQADSAEA 500
Db 516 FTISYDNEKEMTKLDLITSDMA 538

RESULT 11
AAU00629
ID AAU00629 standard; Protein; 586 AA.
XX
AC AAU00629;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #2.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN WO200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
XX
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX
DR WPI; 2001-290917/30.
```


QY 193 EWTHTGOENSWKPKARLKPKPPWAAAFATDEYQWLQIDLNKKKITGIITGTSTWVH 252
 Db 244 ----- 243
 QY 253 NYTVSAIRILYSDGQKWTVPYREPGEVDKIFOGNKDYHQDVRRNFPPIARFIRVNPT 312
 Db 244 -----PGM----- 246
 QY 313 OMQOKIAMKELLGCOFIPKGRPPKLTQPPPPRNSNDLKNTPPKIAKGRAPKFTQPIQ 372
 Db 247 ----- 246
 QY 373 PRSSNEPPAQTEQTATSPDIRNTTVPNTVDVAALVPLVMVLTLLILVCAWH 432
 Db 247 -----NIT-TVAIPSVI--FIALLTGMGFAIC----- 272
 QY 433 RNRKKTKTEGYDLPY-----WDRAGMWMKGKFLPAKAVDHEETP-VRYSSSP--VNHLS 485
 Db 273 --RKRKKKGN--PYVSADAQKTCWKQIKY-----PFAHQSTFTTISYDNE 315
 QY 486 REVTVLQADSAYEAQPLVGGIVGTLHQSTFKP-----EE-----GKEAGYADLDPY 533
 Db 316 KEMTKLDLITSDMADYQQLMIGTGVARKGSTFRPMDTTEEVRVNTEASGH-----Y 370
 QY 534 NSPGQEVYHAYAEPLPITGCVATPII-----MDMSGHTTSVGPSTSTFKATG 583
 Db 371 DCPHRGRHEYALPHTHSEPEATPLVERHLRAHTFTQSGYRVPG-PRPHEHSHSSG 429
 QY 584 NOPPLVGTNTLLRSTSCSAQAQYDTPKAGKPLPAPDELVYQVQSTQEVSGAGRD 643
 Db 430 GPPPATGATQVESYQRPASPCKPVGGYDKP-AASSFLDSRD-----PASQSQMTSGGDD 482
 QY 644 G 644
 Db 483 G 483

RESULT 13
 AAE22715
 ID AAE22715 standard; Protein: 398 AA.
 XX AC AAE22715;
 XX DT 09-AUG-2002 (first entry)
 XX DE Human neuropilin-Hyl protein.
 KW Human; neuropilin-like polypeptide; neuropilin-Hyl; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW myotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW nontropic; neuroprotective; vulnary; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
 KW immunosuppressive; chromosome 6q21.
 OS Homo sapiens.
 XX WO200222815-A1.
 XX PN
 XX PD
 XX PD 21-MAR-2002.
 XX PF 12-SEP-2001; 2001WO-US28488.
 XX PR 11-SEP-2000; 2000US-0659671.
 XX PR 06-SEP-2001; 2001US-317902P.

PA (HYSE-) HYSEQ INC.
 XX Tang YT;
 XX WPI; 2002-393966/42.
 DR N-PSDB; AAD35992.
 XX
 PT Novel isolated human Neuropilin-Hyl and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects
 XX
 PS Claim 3; Page 123-125; 152pp; English.
 XX
 CC The invention relates to neuropilin-like polypeptide (neuropilin-Hyl and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hyl
 CC protein. Neuropilin-Hyl gene is located on chromosome 6q21.
 XX
 SQ Sequence 398 AA;
 Query Match 10.9%; Score 378; DB 23; Length 398;
 Best Local Similarity 29.4%; Pred. No. 7.2e-24;
 Matches 104; Conservative 45; Mismatches 99; Indels 106; Gaps 9;

QY 13 LLELDAGAAQCKYCGGLQNMHSIESKNEITLLFNSGITHVSGRFLASYSVID----- 66
 Db 65 LLFTSSSDQYGPYCG-SMTVPKELLNTSEVTVPFESGSHISGRGFLTYASDHPESQG 123
 QY 67 -----KQDLITCLDTASNEPEFSKYCPAGCLLPFAETSGTIPHGYRDSPLCMAG 118
 Db 124 DRPSEKTLDDQSRFTLAGTCTTFVKDSFS-----TDGT-----SLCKAA 162
 QY 119 VHAGVSVNTLGQISVWISKGIPIYESSIANVTS----- 153
 Db 163 IHAGIADLGGQISVLRKQISRYEGILANGVLSREFFIFRQLFSSVLYFSWGNVHA 222
 QY 154 -----VVGH-----LSTSLFTFTSCYCTGLGME 177
 Db 223 VIELFPHMIVMHSKTRREGSTAEEEGVPKLYLVYIQKQELVQDLVLVATVCGSRSLSE 282
 QY 178 SGVIADPOITASSVLEWTDHTGOENSWKPKARLKPKPPWAAAFATDE-----YQWLQIDL 233
 Db 283 ----PDGQIRASSSWQSVNESGDQVHWSFGQARLQDQGSWASGDSNNHKKPREWLEIDL 338
 QY 234 NKEKITGITGTSTPMVEHNYVVSAYRILYSDGQKWTVPYREPGEVDKIFQGN 287
 Db 339 GEKKKITGITGTSTQSNFNFYVKSVFVNMKNNSKWKTKYKGVNNEKV-RGN 391

RESULT 14

CC modifications in the factor V gene, where (I) expresses an activated
CC protein C (APC) resistant factor V and (II) lacks the ability to express
CC wild-type factor V. The transgenic animals (I) and (II) are useful for
CC screening compounds with anticoagulant activity. Methods from the present
CC invention, and the transgenic animals, are also useful in providing
CC models for human thrombophilia. These models are useful in providing
CC insight into the basic regulatory mechanisms of blood coagulation and
CC pathogenesis of human thrombosis. In addition, factor V null transgenic
CC mice, especially pregnant females may be used as a model system to test
CC in utero gene replacement therapy protocols. The present sequence
CC represents murine factor V, which is used in an example from the present
CC invention.

XX
SQ Sequence 2183 AA;

Query Match 10.2%; Score 355.5; DB 21; Length 2183;
Best Local Similarity 40.1%; Pred. No. 8.1e-21;
Matches 79; Conservative 28; Mismatches 75; Indels 15; Gaps 4;
QY 135 VISKGIPIYESSIANNTSVVGHLSLFTFKTSGCYGLGMSGVADPQITASSVLE- 193
Db 1995 IVARIIRIHPTKSYNPT-----LRLEQGEVNGCSTPLGLEDRIQDKQITASSFKS 2049
QY 194 -WTDHTGQENSWKPKARLKPG--PPNAAFAIDEYQWLQIDLNKEKKITGIITGSTMV 250
Db 2050 WAGDY-----WPSLARLNAQGRVNAWQAKANNKQWLQVDLLKIKVTAIVTQCKSL 2103
QY 251 EHNYYVSAYRILYSDDGQKWTVYRPGVEQDKIFQGNKDYHODVRNNFLPPIIARFIRVN 310
Db 2104 SSEMYYKSYSIQYSDQGVAVKPYROKSSMVDKIFEGNSNTKGHMKNFNPPIISRFRIR 2163
QY 311 PTQWQOKIAKMKELLGC 327
Db 2164 PKTNQSIARLELFGC 2180

Search completed: January 21, 2003, 09:52:42
Job time : 28.0427 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:42:00 ; Search time 27.9573 Seconds
(without alignments)
5372.776 Million cell updates/sec

Title: US-10-060-830-3
Perfect score: 3888
Sequence: 1 MPEFLLLVLLVLLLEDAGA.....TQEVSGAGRDGCDVFKFEL 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_drychea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	3888	100.0	775	4	Q96PD2	Q96pd2 homo sapien
2	3888	100.0	775	4	Q8TDX2	Q8tdx2 homo sapien
3	3415.5	87.8	769	11	Q912V2	Q912v2 rattus norv
4	3382.5	87.0	769	11	Q912V3	Q912v3 mus musculu
5	1396.5	35.9	364	4	Q14089	Q14089 homo sapien
6	1151.5	29.6	251	11	Q9D9K5	Q9d9k5 mus musculu
7	569.5	14.6	503	11	Q9DAJ3	Q9daj3 mus musculu
8	507.5	13.1	460	11	Q9D696	Q9d696 mus musculu
9	412	10.6	432	11	Q8R327	Q8r327 mus musculu
10	402.5	10.4	926	11	Q8Q2V7	Q8q2v7 mus musculu
11	400.5	10.3	609	4	Q96I90	Q96i90 homo sapien
12	400.5	10.3	644	4	Q96I95	Q96ih5 homo sapien
13	400.5	10.3	704	4	Q9H2E1	Q9h2e1 homo sapien
14	395.5	10.2	901	4	Q9H2E4	Q9h2e4 homo sapien
15	395.5	10.2	901	4	Q9H2D5	Q9h2d5 homo sapien
16	395.5	10.2	906	4	Q9H2E3	Q9h2e3 homo sapien

17	395.5	10.2	906	4	Q9H2D4	Q9h2d4 homo sapien
18	392	10.1	919	13	Q8UVRO	Q8uvr0 gallus gall
19	392	10.1	936	13	Q8UVQ9	Q8uvq9 gallus gall
20	386	9.9	555	4	Q9H2E2	Q9h2e2 homo sapien
21	383.5	9.9	923	13	Q8QFX6	Q8qfx6 brachydanio
22	382.5	9.8	921	11	Q9QX38	Q9qx38 rattus norv
23	359	9.2	2119	13	Q90X47	Q90x47 brachydanio
24	355.5	9.1	2183	11	O88783	O88783 mus musculu
25	355	9.1	480	4	O43854	O43854 homo sapien
26	352	9.1	480	11	O35474	O35474 mus musculu
27	351	9.0	2224	4	O43737	O43737 homo sapien
28	343.5	8.8	779	4	Q9NTT3	Q9ntt3 homo sapien
29	327.5	8.4	2343	6	O18806	O18806 canis fami
30	321.5	8.3	426	11	Q9WT53	Q9wt53 mus musculu
31	321.5	8.3	2343	6	O62730	O62730 canis fami
32	321	8.3	463	11	Q9RIX9	Q9rix9 mus musculu
33	308.5	7.9	216	4	O14286	O14286 homo sapien
34	271	7.0	363	6	O77718	O77718 equus cabal
35	269.5	6.9	224	11	Q9RIM6	Q9rim6 mus musculu
36	267.5	6.9	858	5	O76470	O76470 lytechinus
37	239	6.1	3620	6	Q9TU53	Q9tu53 canis fami
38	228	5.9	335	4	Q9BTU9	Q9bt19 homo sapien
39	227.5	5.9	1128	11	O88442	O88442 mus musculu
40	224.5	5.8	1158	4	Q14113	Q14113 homo sapien
41	220.5	5.7	764	11	Q9D2L5	Q9d2l5 mus musculu
42	220.5	5.7	764	11	O54860	O54860 mus musculu
43	220.5	5.7	764	11	O8VDQ4	O8vdq4 mus musculu
44	216	5.6	1524	13	Q91674	Q91674 xenopus lae
45	214.5	5.5	3623	11	O70244	O70244 rattus norv

ALIGNMENTS

RESULT 1

Q96PD2	PRELIMINARY;	PRT;	775 AA.
ID	Q96PD2;		
AC	Q96PD2;		
DC	01-DEC-2001 (Tremblrel. 19, Created)		
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Endothelial and smooth muscle cell-derived neuropilin-like protein.		
DE	protein.		
GN	ESDN.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	PubMed=11447234;		
RA	Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,		
RA	Matsumori A., Sasayama S., Honjo T., Tashiro K.;		
RT	"ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular		
RT	Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is		
RT	Up-regulated after Vascular Injury.";		
RL	J. Biol. Chem. 276:34105-34114(2001).		
CC	-1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL		
CC	REGULATORS.		
CC	EMBL; AF387547; AAL30179.1; .		
DR	InterPro; IPR000859; CUB_domain.		
DR	InterPro; IPR000421; FA58.C.		
DR	InterPro; IPR004043; LCCL_dom.		
DR	Pfam; PF00431; CUB; 1.		
DR	Pfam; PF00754; FS_F8_type_C; 1.		
DR	PROSITE; PS01180; CUB; 1.		
KW	DNA-binding; Transcription regulation.		
SQ	SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;		

Query Match 100.0%; Score 3888; DB 4; Length 775;
Best Local Similarity 100.0%; Pred. No. 3.7e-304;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MFLFLLLVLLVLLLEDAAGGCGGHTVILGPESGTLTISINYPTPNSTVCEWEIRVK 60
DB 47 MFLFLLLVLLVLLLEDAAGGCGGHTVILGPESGTLTISINYPTPNSTVCEWEIRVK 106
QY 61 MGERVRIKFGDFDIEDSDSCHFNLYRIYNGVSRTEIGKYCGGLQMNHIESKGNIEIT 120
DB 107 MGERVRIKFGDFDIEDSDSCHFNLYRIYNGVSRTEIGKYCGGLQMNHIESKGNIEIT 166
QY 121 LFMGSIHVSGRGLFASVSDKODLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 180
DB 167 LFMGSIHVSGRGLFASVSDKODLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 226
QY 181 PHGYRDSPLCMAGVHAGVWSNTLGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFT 240
DB 227 PHGYRDSPLCMAGVHAGVWSNTLGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFT 286
QY 241 FKTSCYCTLGMESGVADPOITASSVLEWTDHTGOENSWPKPKARLKKPGPPWAAAFATD 300
DB 287 FKTSCYCTLGMESGVADPOITASSVLEWTDHTGOENSWPKPKARLKKPGPPWAAAFATD 346
QY 301 EYQWLQIDLNEKKITGIIITGSTMVHNYSAYRILYSDGOKWTYVREPVEQDKIF 360
DB 347 EYQWLQIDLNEKKITGIIITGSTMVHNYSAYRILYSDGOKWTYVREPVEQDKIF 406
QY 361 QGNKDYHQDVRRNLPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPPKLTQPPPP 420
DB 407 QGNKDYHQDVRRNLPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPPKLTQPPPP 466
QY 421 RNSNDLKNTTAPPKIAGRAPKFTQPLQPRSSNEFFPAQTEOTTASPDIRNTVTPNVTKD 480
DB 467 RNSNDLKNTTAPPKIAGRAPKFTQPLQPRSSNEFFPAQTEOTTASPDIRNTVTPNVTKD 526
QY 481 VALAAVLPVLMVLTLLILVCAWHNRNKKTEGYDLPYWDRAWGMKQFLPAK 540
DB 527 VALAAVLPVLMVLTLLILVCAWHNRNKKTEGYDLPYWDRAWGMKQFLPAK 586
QY 541 AVDHEETPVRYSSSEVNHLSPREVTTLQADSAYEAQPLVGGIVGTLHORSTFKPEEGKE 600
DB 587 AVDHEETPVRYSSSEVNHLSPREVTTLQADSAYEAQPLVGGIVGTLHORSTFKPEEGKE 646
QY 601 AGYADLDYPNPGQEVYHAYAEPLDITGPEYATPIIMDSGHPTTSVGQPSSTFKATGN 660
DB 647 AGYADLDYPNPGQEVYHAYAEPLDITGPEYATPIIMDSGHPTTSVGQPSSTFKATGN 706
QY 661 OPPPLVGYNTLLSRTDSCSSAQAYDTPKAGKPGCLPAPDELVYQVPOSTQEVSGAGRDG 720
DB 707 OPPPLVGYNTLLSRTDSCSSAQAYDTPKAGKPGCLPAPDELVYQVPOSTQEVSGAGRDG 766
QY 721 ECDVFEKIL 729
DB 767 ECDVFEKIL 775

RESULT 2
Q8TDX2 PRELIMINARY; PRT; 775 AA.
ID Q8TDX2
AC Q8TDX2;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE CLCP1.
GN CLCP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Koshikawa K., Osada H., Kozaki K., Konishi H., Masuda A.,
RA Tatematsu Y., Mitsudomi T., Nakao A., Takahashi T.;
RT "Significant up-regulation of a novel gene, CLCP1, in a highly
RT metastatic lung cancer subline as well as in lung cancers in vivo.";
```

```

RL Oncogene 0:0-0(2002).
DR EMBL; AB073146; BAB91138.1; -.
SQ SEQUENCE 775 AA; 85073 MW; 6D07223B21BE5A42 CRC64;

Query Match 100.0%; Score 3888; DB 4; Length 775;
Best Local Similarity 100.0%; Pred. No. 3,7e-304;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFLFLLLVLLVLLLEDAAGGCGGHTVILGPESGTLTISINYPTPNSTVCEWEIRVK 60
DB 47 MFLFLLLVLLVLLLEDAAGGCGGHTVILGPESGTLTISINYPTPNSTVCEWEIRVK 106
QY 61 MGERVRIKFGDFDIEDSDSCHFNLYRIYNGVSRTEIGKYCGGLQMNHIESKGNIEIT 120
DB 107 MGERVRIKFGDFDIEDSDSCHFNLYRIYNGVSRTEIGKYCGGLQMNHIESKGNIEIT 166
QY 121 LFMGSIHVSGRGLFASVSDKODLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 180
DB 167 LFMGSIHVSGRGLFASVSDKODLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 226
QY 181 PHGYRDSPLCMAGVHAGVWSNTLGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFT 240
DB 227 PHGYRDSPLCMAGVHAGVWSNTLGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFT 286
QY 241 FKTSCYCTLGMESGVADPOITASSVLEWTDHTGOENSWPKPKARLKKPGPPWAAAFATD 300
DB 287 FKTSCYCTLGMESGVADPOITASSVLEWTDHTGOENSWPKPKARLKKPGPPWAAAFATD 346
QY 301 EYQWLQIDLNEKKITGIIITGSTMVHNYSAYRILYSDGOKWTYVREPVEQDKIF 360
DB 347 EYQWLQIDLNEKKITGIIITGSTMVHNYSAYRILYSDGOKWTYVREPVEQDKIF 406
QY 361 QGNKDYHQDVRRNLPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPPKLTQPPPP 420
DB 407 QGNKDYHQDVRRNLPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPPKLTQPPPP 466
QY 421 RNSNDLKNTTAPPKIAGRAPKFTQPLQPRSSNEFFPAQTEOTTASPDIRNTVTPNVTKD 480
DB 467 RNSNDLKNTTAPPKIAGRAPKFTQPLQPRSSNEFFPAQTEOTTASPDIRNTVTPNVTKD 526
QY 481 VALAAVLPVLMVLTLLILVCAWHNRNKKTEGYDLPYWDRAWGMKQFLPAK 540
DB 527 VALAAVLPVLMVLTLLILVCAWHNRNKKTEGYDLPYWDRAWGMKQFLPAK 586
QY 541 AVDHEETPVRYSSSEVNHLSPREVTTLQADSAYEAQPLVGGIVGTLHORSTFKPEEGKE 600
DB 587 AVDHEETPVRYSSSEVNHLSPREVTTLQADSAYEAQPLVGGIVGTLHORSTFKPEEGKE 646
QY 601 AGYADLDYPNPGQEVYHAYAEPLDITGPEYATPIIMDSGHPTTSVGQPSSTFKATGN 660
DB 647 AGYADLDYPNPGQEVYHAYAEPLDITGPEYATPIIMDSGHPTTSVGQPSSTFKATGN 706
QY 661 OPPPLVGYNTLLSRTDSCSSAQAYDTPKAGKPGCLPAPDELVYQVPOSTQEVSGAGRDG 720
DB 707 OPPPLVGYNTLLSRTDSCSSAQAYDTPKAGKPGCLPAPDELVYQVPOSTQEVSGAGRDG 766
QY 721 ECDVFEKIL 729
DB 767 ECDVFEKIL 775

RESULT 3
Q912V2 PRELIMINARY; PRT; 769 AA.
ID Q912V2
AC Q912V2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE CLCP1.
GN CLCP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX PubMed=11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury.";
RL J. Biol. Chem. 276:34105-34114(2001).
CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AF387549; AAL30180.1; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR PROSITE; PS01180; CUB; 1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 769 AA; 83915 MW; 5B7936C8FA063A93 CRC64;

Query Match 87.88; Score 3415.5; DB 11; Length 769;
Best Local Similarity 87.7%; Pred. No. 4e-266;
Matches 634; Conservative 40; Mismatches 46; Indels 3; Gaps 2;

QY 7 LLLVLLLEDAGAGCGGCGHTVLGPGESGLTISINYPQTPNSVCEWIRVKMGVERV 66
DB LLLVLLLEDAGAGCGGCGHTVLGPGESGLTISINYPQTPNSVCEWIRVKMGVERV 66
QY 50 LLLVLLLEDAGAGCGGCGHTVLGPGESGLTISINYPQTPNSVCEWIRVKMGVERV 109
DB LLLVLLLEDAGAGCGGCGHTVLGPGESGLTISINYPQTPNSVCEWIRVKMGVERV 109
QY 67 IKFGDFIEDSDSCHFNRYLRIYNGIGVSRTEIGKYGGLQMNHSIESKGNIEITLLFMSG 126
DB IKFGDFIEDSDSCHFNRYLRIYNGIGVSRTEIGKYGGLQMNHSIESKGNIEITLLFMSG 126
QY 110 IKFGDFIEDSDSCHFNRYLRIYNGIGVSRTEIGKYGGLQMNHSIESKGNIEITLLFMSG 169
DB IKFGDFIEDSDSCHFNRYLRIYNGIGVSRTEIGKYGGLQMNHSIESKGNIEITLLFMSG 169
QY 127 IHVSGRGLASYSVIDKODLITCLDTASNFLPEPESKYPACGLLPFAEISGTPHGYRD 186
DB IHVSGRGLASYSVIDKODLITCLDTASNFLPEPESKYPACGLLPFAEISGTPHGYRD 186
QY 170 THAAGRGFLASYSVIDKEDLITCLDTASNFLPEPESKYPACGLLPFAEISGTPHGYRD 229
DB THAAGRGFLASYSVIDKEDLITCLDTASNFLPEPESKYPACGLLPFAEISGTPHGYRD 229
QY 187 SSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVVGHLSLSTLFTKTSKC 246
DB SSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVVGHLSLSTLFTKTSKC 246
QY 230 SSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVVGHLSLSTLFTKTSKC 289
DB SSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVVGHLSLSTLFTKTSKC 289
QY 247 YGTLMGESVIADPOITASSVLEWTDHTGQENSWPKKARLKKPGPPWAAAFATDEYQWLQ 306
DB YGTLMGESVIADPOITASSVLEWTDHTGQENSWPKKARLKKPGPPWAAAFATDEYQWLQ 306
QY 290 YGTLMGESVIADPOITASSVLEWTDHTGQENSWPKKARLKKPGPPWAAAFATDEYQWLQ 349
DB YGTLMGESVIADPOITASSVLEWTDHTGQENSWPKKARLKKPGPPWAAAFATDEYQWLQ 349
QY 307 IDLNKEKKITGITTTGSTMVHEHNYVSAYRILYSDGQKWTYRPGVEQDKIFQGNKDY 366
DB IDLNKEKKITGITTTGSTMVHEHNYVSAYRILYSDGQKWTYRPGVEQDKIFQGNKDY 366
QY 350 IDLNKEKKITGITTTGSTMVHEHNYVSAYRILYSDGQKWTYRPGVEQDKIFQGNKDY 409
DB IDLNKEKKITGITTTGSTMVHEHNYVSAYRILYSDGQKWTYRPGVEQDKIFQGNKDY 409
```

```
DB 708 GTNTLLSRDSCSSGQAQYDTKGGKPA-AAPEELVYVQPSTQEAAGAGRDEKFDAPK 766
QY 727 EIL 729
DB 767 ETL 769

RESULT 4
Q91ZV3 PRELIMINARY; PRT; 769 AA.
ID Q91ZV3;
AC Q91ZV3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like
DE protein.
DS ESDN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR;
RX PubMed=11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury.";
RL J. Biol. Chem. 276:34105-34114(2001).
CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AF387548; AAL30179.1; -.
DR MGD; MGI:1920629; ESDN.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR PROSITE; PS01180; CUB; 1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;

Query Match 87.0%; Score 3382.5; DB 11; Length 769;
Best Local Similarity 86.7%; Pred. No. 1.8e-263;
Matches 627; Conservative 41; Mismatches 52; Indels 3; Gaps 2;

QY 7 LLLVLLLEDAGAGCGGCGHTVLGPGESGLTISINYPQTPNSVCEWIRVKMGVERV 66
DB LLLVLLLEDAGAGCGGCGHTVLGPGESGLTISINYPQTPNSVCEWIRVKMGVERV 66
QY 50 LLLVLLLEDAGAGCGGCGHTVLGPGESGLTISINYPQTPNSVCEWIRVKMGVERV 109
DB LLLVLLLEDAGAGCGGCGHTVLGPGESGLTISINYPQTPNSVCEWIRVKMGVERV 109
QY 67 IKFGDFIEDSDSCHFNRYLRIYNGIGVSRTEIGKYGGLQMNHSIESKGNIEITLLFMSG 126
DB IKFGDFIEDSDSCHFNRYLRIYNGIGVSRTEIGKYGGLQMNHSIESKGNIEITLLFMSG 126
QY 110 IKFGDFIEDSDSCHFNRYLRIYNGIGVSRTEIGKYGGLQMNHSIESKGNIEITLLFMSG 169
DB IKFGDFIEDSDSCHFNRYLRIYNGIGVSRTEIGKYGGLQMNHSIESKGNIEITLLFMSG 169
QY 127 IHVSGRGLASYSVIDKODLITCLDTASNFLPEPESKYPACGLLPFAEISGTPHGYRD 186
DB IHVSGRGLASYSVIDKODLITCLDTASNFLPEPESKYPACGLLPFAEISGTPHGYRD 186
QY 170 THAAGRGFLASYSVIDKEDLITCLDTASNFLPEPESKYPACGLLPFAEISGTPHGYRD 229
DB THAAGRGFLASYSVIDKEDLITCLDTASNFLPEPESKYPACGLLPFAEISGTPHGYRD 229
QY 187 SSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVVGHLSLSTLFTKTSKC 246
DB SSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVVGHLSLSTLFTKTSKC 246
QY 230 SSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVVGHLSLSTLFTKTSKC 289
DB SSPLCMAGVHAGVSVNTLGGQISVVISKIPYESSLANNVTSVVGHLSLSTLFTKTSKC 289
QY 247 YGTLMGESVIADPOITASSVLEWTDHTGQENSWPKKARLKKPGPPWAAAFATDEYQWLQ 306
DB YGTLMGESVIADPOITASSVLEWTDHTGQENSWPKKARLKKPGPPWAAAFATDEYQWLQ 306
QY 290 YGTLMGESVIADPOITASSVLEWTDHTGQENSWPKKARLKKPGPPWAAAFATDEYQWLQ 349
DB YGTLMGESVIADPOITASSVLEWTDHTGQENSWPKKARLKKPGPPWAAAFATDEYQWLQ 349
QY 307 IDLNKEKKITGITTTGSTMVHEHNYVSAYRILYSDGQKWTYRPGVEQDKIFQGNKDY 366
DB IDLNKEKKITGITTTGSTMVHEHNYVSAYRILYSDGQKWTYRPGVEQDKIFQGNKDY 366
QY 350 IDLNKEKKITGITTTGSTMVHEHNYVSAYRILYSDGQKWTYRPGVEQDKIFQGNKDY 409
DB IDLNKEKKITGITTTGSTMVHEHNYVSAYRILYSDGQKWTYRPGVEQDKIFQGNKDY 409
```


OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ILS, AND ISS;
RX MEDLINE=21363810; PubMed=11471062;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RT "High-throughput sequence identification of gene coding variants
RT within alcohol-related QTLs";
RL Mamm. Genome 12:657-663(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ILS, AND ISS;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF483506; AAL90780.1; -;
DR EMBL; AF483507; AAL90781.1; -;
SQ SEQUENCE 926 AA; 104054 MW; 8073D6478C9A8467 CRC64;

Query Match 10.4%; Score 402.5; DB 11; Length 926;
Best Local Similarity 24.1%; Pred. No. 2.3e-23;
Matches 165; Conservative 90; Mismatches 290; Indels 139; Gaps 26;

Qy 1 MPFLLLLLLLLLL--EDAGAQGGCGHVLGPESGTLTSINYPQYPSNVCWEWR 58
Db 1 MDMPLTWVFLALYFSGHEVRSQDPCCGRSLNLSADGIITSPGYQDPYSHQNCWEIY 60

Qy 59 V-KMGERVRIKFG-DFDIEDSDSCHFNRLIYNGIGVSRTEIGKYCGGLGLOWN---HSIE 113
Db 61 APEPNQKIVLNFPHFELEKID-CYDFIEIRDGDSADLLGKHC-----NIAPPII 114

Qy 114 SKGNEITLFLMSGHVGSGFLASVIDKQDLITCLPTASNLPEFESKYCPACGLPFF 173
Db 115 SSGSVLYIKFTSDYARQAGFSLRIEIKTGS----EDCSKNFTSPNGTIESPG----F 165

Qy 174 AEISTGPHGVYDSDPLCMAGVAGVSNLTGGQIS-----VVISKGP- 217
Db 166 PE---KYPHNL-DCTFTTLAKRMEIILQFTFLEHDPLOVGECDKYDWDLDWDGIPH 221

Qy 218 -----YYESSLANNVSVVGHLSLSTLFTKTS-----GCV 247
Db 222 VGPLGKCGTKPKSLRSSGILS---LTFHTDMAVAKDGFSAHYLIHQEPENFOCN 278

Qy 248 GPLMGESVIADPQITASSVLEWTDHTGOENSWKPKKARLKPFPWAAFAATDEYQWLQI 307
Db 279 VPLMGESRIANEQISASS-----TFSGGRWTQQSRLHGDGNGWTPNLDSNKEYLQV 331

Qy 308 DLNKEKKITGITTG--STMVHNNYVSAIRILYSDDGOKWTVYREPGEQDKIFQGNKD 365
Db 332 DLRFUTMTATQCAISRETQKGYVKSYLEVSTNGEDMWVYRHG---KNHKIFQAND 389

Qy 366 YHQDVRRNPLPPIARFIRVNPQOQKIAMKMLLGCQFIPKRPKLTQPPPRNSND 425
Db 390 RTEVVLNKLHPLLTFRIRPQTHLGLALRLFLGC-----RVTDAPCSNMLGM 440

Qy 426 LKNTAPKIAKRAPK-----TQPLQPRSSNEPFAQTEQTASPDINTVTNVTNKD 480
Db 441 LSGLIADTQISASSSTREVLWSPSAARLVSSRSRGWFPNPPQAQGEWLQDGLGTPKTVKG 500

Qy 481 V-----ALAAVLVPVLMVLTLLILYCAHHW-----RNRKKTEGT--VDLPY 524
Db 501 VTIQARGGDSITAYEARAFVRFKYSVSLNGKDMKEYIODPRTQQTCLKFEGNNHYDTP--- 558

Qy 525 DRAGWKGKQFLPAKVDHEETPVYSSSEVNHLSPREVTTVLOADSAEY--AOPLVGG 582
Db 559 -----DIRREDP-----VPAQYVRVYFERWSPAGIMGRLEVLGCDWTSKPTVET 603

Qy 583 IVGTLHQSTFKP-----EGKEAG 602

Db 604 LGPTVKSEETTPYPMDEDATECG 627

RESULT 11
Q96190 PRELIMINARY; PRT; 609 AA.
AC Q96190;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to neuropilin 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007737; AAH07737.1; -;
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C.1; UNKNOWN_2.
DR PROSITE; PS01286; FA58C.2; UNKNOWN_2.
SQ SEQUENCE 609 AA; 68347 MW; 798AAFD2D568C127 CRC64;

Query Match 10.3%; Score 400.5; DB 4; Length 609;
Best Local Similarity 28.1%; Pred. No. 1.7e-23;
Matches 130; Conservative 66; Mismatches 178; Indels 89; Gaps 18;

Qy 6 LLLVLLVLLLDAGAAQGGCGHVTVLGPESGTLTSINYPQYPSNVCWEWRV-KMGER 64
Db 7 LLCALVALVAPAGAFNRDKCGDTIKIESPGYLTSPGYPHSYHPSEKCEWLIIQADPYQR 66

Qy 65 VRIKFG-DFDIEDSDSCHFNRLIYNGIGVSRTEIGKYCGGLGLOWNHSTESKGNITLLF 123
Db 67 IMINENPHFLEDRD-CKYDYVEFEDGENENGHFRGKFC--GKIAPPVWSSGPFLEKIF 123

Qy 124 MSGIHVSGRGFLASYSVIDK-----QDLIT----- 148
Db 124 VSDYETHGAGFSIRYEIFKRGPECSQNVTPSGVIKSPGPEKYPNSLECTYIVFAPKMS 183

Qy 149 --CLDTASNLFEPESKYCPACGLLPFAEISCTIPHGYDSSPLCMAGVHAGVVS---N 202
Db 184 EILFEFSFDLEPDSNP--PGMFCRYDRLE--IWDGFPDVGPP--HIGYCCQKTPGRIR 237

Qy 203 TLGGQISVVISKIGIPYESSLAN-----NVTSVVGHLSLSTLFTKTSQCYGLMGESGVI 257
Db 238 SSSGILSMVF-----YTDSAIKAGFSANYSVLQSSVSED---FK--CMEALGMESGFI 286

Qy 258 ADPQITASSVLEWTDHTGOENSWKPKKARLKPFPWAAFAATDEYQWLQIDLNKEKKITG 317
Db 287 HSDQITASS-----QYSTNWSAERSRLNYPENGWTPGSDSYREWIQVDLGLRFEVTA 338

Qy 318 IITTG--STMVHNNYVSAIRILYSDDGOKWTVYREPGEQDKIFQGNKDYHQDYRRNNEF 375
Db 339 VGTQGAISRETKKYVTKYKIDVSSNGEDWITIKEG--NKPVFCQGNTPDVTVDVVAVPP 396

Qy 376 PPIIARFIRVNPQOQKIAMKMLLGCQFIPKRPKLTQPP 418
Db 397 KPLIRFRIKATWETGISMRFEVYGC-----KITDYP 430

RESULT 12
Q961H5 PRELIMINARY; PRT; 644 AA.
AC Q961H5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

```
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Neuropilin 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007533; AAH07533.1; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C.1; 1.
DR PROSITE; PS01286; FA58C.2; UNKNOWN_2.
SQ SEQUENCE 644 AA; 71907 MW; 9204B6F5C3CD2A21 CRC64;

Query Match 10.3%; Score 400.5; DB 4; Length 644;
Best Local Similarity 28.1%; Pred. No. 1.9e-23;
Matches 130; Conservative 66; Mismatches 178; Indels 89; Gaps 18;

QY 6 LLLLVLLLLLEDAGAOQGGCGHTVLPESGTLTSINYPQTVNSTVCEWEIRV-KMGER 64
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 65 VRIKFG-DFDIEDSDSCHFNLYRIYNGIGVSRTEIKYCGGLQMNHSIESKNEITLIF 123
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 67 IMINENPHFDLEDRD-CKYDYVEVDFGENENHFRGKFC--GKIAPPPVSSGPFLEIKF 123
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 124 MSGIHVSGRGLASVSDIK-----QDLIT----- 148
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 124 VSDYETHGAGFIRYEIFKRGPECSQNTTSPGVKSPGFPPEKYPNSLECTYIVFAPKMS 183
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 149 --CLDTASNFLEPEFSKYCPAGCLLPFAEISGTIPHGVRDSSPLCMAGVHAGVVS---N 202
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 184 EIILEFESFDLPDSNP--PGMFCRYDLE--INDGFPDVP--HIGRYCQKTPGRIR 237
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 203 TLGGQISVVISKGIPIYESSLAN-----NVTSVVGHLSLTFKTSYCYGLTGMESGYI 257
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 238 SSSGILSMVF-----YTDSAIAKEGFSANYSVLQSSVED---FK--CMEALGMESGEI 286
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 258 ADPQITASSVLEWTDHTGOENSWKPKARKLKPFPWAAFAFATDEYQWLQIDLNKEKITG 317
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 287 HSDQITASS-----QYSTNWSAERSRLNYPENGWTPGEDSYREWIQVLDGLLRFTVA 338
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 318 IITTG--STMVEHNYVSAIRILYSDDGOKVTYREPVEQDKIFOGNKDYHODVRRNLF 375
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 339 VGTQGAISKETKKKYVTKYKIDVSNNGEDWITIEG--NKPVLFGQNTPTDVVVAVFP 396
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 376 PPIARFIRVNPQOQKIAMKMLGCOFIPKGRPPKLTQPP 418
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 397 KPLITRFVRIKPAWTGISMREFYVGC-----KITDYP 430
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13
Q9H2E1 PRELIMINARY; PRT; 704 AA.
AC Q9H2E1;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Neuropilin-1 soluble isoform 11.
GN NRPI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC TISSUE=PROSTATIC ADENOCARCINOMA;
EX MEDLINE=20564205; PubMed=11112349;
RA Rossignol M.; Gagnon M.B.; Klagsbrun M.;
RT "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms."
RL Genomics 70:211-222(2000).
CC -|- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DR EMBL; AF280547; AAG41406.1; -.
DR HSSP; P12259; IC2T.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C.1; 1.
DR PROSITE; PS01286; FA58C.2; 2.
SQ SEQUENCE 704 AA; 79002 MW; 1625DDDD561459307 CRC64;

Query Match 10.3%; Score 400.5; DB 4; Length 704;
Best Local Similarity 28.1%; Pred. No. 2.2e-23;
Matches 130; Conservative 66; Mismatches 178; Indels 89; Gaps 18;

QY 6 LLLLVLLLLLEDAGAOQGGCGHTVLPESGTLTSINYPQTVNSTVCEWEIRV-KMGER 64
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 65 VRIKFG-DFDIEDSDSCHFNLYRIYNGIGVSRTEIKYCGGLQMNHSIESKNEITLIF 123
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 67 IMINENPHFDLEDRD-CKYDYVEVDFGENENHFRGKFC--GKIAPPPVSSGPFLEIKF 123
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 124 MSGIHVSGRGLASVSDIK-----QDLIT----- 148
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 124 VSDYETHGAGFIRYEIFKRGPECSQNTTSPGVKSPGFPPEKYPNSLECTYIVFAPKMS 183
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 149 --CLDTASNFLEPEFSKYCPAGCLLPFAEISGTIPHGVRDSSPLCMAGVHAGVVS---N 202
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 184 EIILEFESFDLPDSNP--PGMFCRYDLE--INDGFPDVP--HIGRYCQKTPGRIR 237
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 203 TLGGQISVVISKGIPIYESSLAN-----NVTSVVGHLSLTFKTSYCYGLTGMESGYI 257
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 238 SSSGILSMVF-----YTDSAIAKEGFSANYSVLQSSVED---FK--CMEALGMESGEI 286
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 258 ADPQITASSVLEWTDHTGOENSWKPKARKLKPFPWAAFAFATDEYQWLQIDLNKEKITG 317
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 287 HSDQITASS-----QYSTNWSAERSRLNYPENGWTPGEDSYREWIQVLDGLLRFTVA 338
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 318 IITTG--STMVEHNYVSAIRILYSDDGOKVTYREPVEQDKIFOGNKDYHODVRRNLF 375
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 339 VGTQGAISKETKKKYVTKYKIDVSNNGEDWITIEG--NKPVLFGQNTPTDVVVAVFP 396
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 376 PPIARFIRVNPQOQKIAMKMLGCOFIPKGRPPKLTQPP 418
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 397 KPLITRFVRIKPAWTGISMREFYVGC-----KITDYP 430
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 14
Q9H2E4 PRELIMINARY; PRT; 901 AA.
AC Q9H2E4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Neuropilin-2b(O).
GN NRPI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RX MEDLINE=20564205; PubMed=11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms.";
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL: AF280544; AA41403.1; -.
DR HSP; P12259; 1CZT.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00431; CUB_2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; UNKNOWN_1.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS50060; MAM_2; 1.
DR GlycoProtein; Neurone; Receptor; Transmembrane.
KW GPCROTEIN; NEURONE; RECEPTOR; TRANSMEMBRANE.
SQ SEQUENCE 901 AA; 101380 MW; 1F3107A9CC665EA2 CRC64;

Query Match 10.2%; Score 395.5; DB 4; Length 901;
Best Local Similarity 23.6%; Pred. No. 8e-23;
Matches 161; Conservative 92; Mismatches 295; Indels 133; Gaps 25;

Qy 1 MPFL--LLLVLLLLLEDAGAQGGCGHTVLPESGTLTSINYPQYPSNSTVCEWIR 58
Db 1 MDPEPLTWVFLALYFSRHQVRGQPPCGGRLNKSDAGYITSPGYDPYSHQNCWIVY 60

Qy 59 V-KMGERVRIKFG-DFDIEDSDSCHFNRLRYNGVIGVSRTEIGKYCGGLQWN---HSIE 113
Db 61 APEPNQKIVLNFNPFIEKH-D-KYDFIEIRDGSDSADLLGKHG-----NTAPPII 114

Qy 114 SKGNEITLLFMSGIHVSGRGLASYVIDKODLTICLDTASNFLPEFSKYCPAGCLLPF 173
Db 115 SSGSMLYIKFTSDYARQAGFSLRYEIKTGS-----EDCSKNFTSPNGTIESPG---F 165

Qy 174 AEISGTIPGHVYRDSPLCMAGHAGVSVNTLGGQIS-----VVISKGIP- 217
Db 166 PE---KYPHNL-DCTFTILAKPKMEIILQFLIFDLEHDPLOVGECDCKYDWDIDWGI 221

Qy 218 -----YYESSLANNVTSVVGHLSTSLFT-----FRTSGCYGTL 250
Db 222 VGPLIGKYGKTKPSELSSSTGILSLFHTDMYAKDGFSAHYLVHQEPLNFQCNVPL 281

Qy 251 GMESGVADPOITASSVLEWTDHTGQENSWKPKARLKKPGPPWAAFAFDEYQWLQIDLN 310
Db 282 GMESGRITANEQISASS-----TVSDGRWTPQQSRHLGDDNGWTPNLDNSKEYLQVDLR 334

Qy 311 KEKITGIITG--STWVEHNYYSAYRILISDDGQKTVYREPVQEDKIFQGNKDYHQ 368
Db 335 FLTMLTATATGAISSRETQNGYYKSYKLEYSTNGEDWMVYRHG--KNHKVFQANNDATE 392

Qy 369 DVRRNFLPITARFIRVNPTQWQOKIAMKMLLGCQFIPKGRPPKLTOPPPPPRNSNDLKN 428
Db 393 VVLNKLHAPLITRVRIRPQWHSIALRLEFG-----RYTDAPCSNMLGLMSG 443

Qy 429 TTAPPKIAKRAPK-----TQPLQPRSSNEFFPAQTEOTASDIRNTVTVPNTKDV-- 481
Db 444 LIADSOIASSTQEVLMSPSARLVSSRSNGFPRIPQAPQGEWLOVDLGTPTKTVGVII 503

Qy 482 -----ALAAVYLVVYVWLTTLILVCAWH-----RNKKKTGEGT--YDLPYWDRA 527
Db 504 QGARGGDSITAVEARAFVRKFKVSYSLNGKDWYIQDPRTOQPKLFEGNHYDTP----- 558
```

```

Qy 528 GWMKGMKQFLPAKAVDHEETPVRYSSSVSNHLSPREVTVLQADSAEY--AQPLVGIVG 585
Db 559 ----DIRRFDP-----IPAQYRVYVPERWSPAGIMRLEVLCGDMTDSKPTVKTLGP 606

Qy 586 TLHORSTFKP-----EGKEAG 602
Db 607 TVKSEETTTPTTEATECG 627

RESULT 15
Q9H2D5 ID Q9H2D5 PRELIMINARY; PRT; 901 AA.
AC Q9H2D5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Neuropilin-2b(0).
GN NRP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564205; PubMed=11112349;
RA Rossignol M., Gagnon M.L., Klagsbrun M.;
RT "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT Identification and Distribution of Splice Variants and Soluble
RT Isoforms.";
RL Genomics 70:211-222(2000).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL: AF281074; AA41899.1; -.
DR HSP; P12259; 1CZT.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF00629; MAM; 1.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; UNKNOWN_1.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS50060; MAM_2; 1.
DR GlycoProtein; Neurone; Receptor; Transmembrane.
KW GPCROTEIN; NEURONE; RECEPTOR; TRANSMEMBRANE.
SQ SEQUENCE 901 AA; 101381 MW; AL310304AFD086AB CRC64;
```

Query Match 10.2%; Score 395.5; DB 4; Length 901;
Best Local Similarity 23.6%; Pred. No. 8e-23;
Matches 161; Conservative 92; Mismatches 295; Indels 133; Gaps 25;

```

Qy 1 MPFL--LLLVLLLLLEDAGAQGGCGHTVLPESGTLTSINYPQYPSNSTVCEWIR 58
Db 1 MDPEPLTWVFLALYFSRHQVRGQPPCGGRLNKSDAGYITSPGYDPYSHQNCWIVY 60

Qy 59 V-KMGERVRIKFG-DFDIEDSDSCHFNRLRYNGVIGVSRTEIGKYCGGLQWN---HSIE 113
Db 61 APEPNQKIVLNFNPFIEKH-D-KYDFIEIRDGSDSADLLGKHG-----NTAPPII 114

Qy 114 SKGNEITLLFMSGIHVSGRGLASYVIDKODLTICLDTASNFLPEFSKYCPAGCLLPF 173
Db 115 SSGSMLYIKFTSDYARQAGFSLRYEIKTGS-----EDCSKNFTSPNGTIESPG---F 165

Qy 174 AEISGTIPGHVYRDSPLCMAGHAGVSVNTLGGQIS-----VVISKGIP- 217
Db 166 PE---KYPHNL-DCTFTILAKPKMEIILQFLIFDLEHDPLOVGECDCKYDWDIDWGI 221
```

QY	218	-----YYESSLANNVTSVVGHLSLSLTFT-----FKTSCYGTGL	250
		: : : : : : : : : : : : :	
Db	222	VGPLIGKYGCTKTPTSELRSSTGLSUTFTHTMAVARDKGFSARYLYLVHOBEPLENFQCNPVL	281
		: : : : : : : : : : : : :	
QY	251	GMSGVIAIDPOITASSVLEWTDHTGOENSWKKARLKPGPPWAAFATDEYQWLQIDLN	310
		: : : : : : : : : : : : :	
Db	282	GMSGRIANEQISASS-----TYSDBGRTWQQSRHLGGDNGWTPLNDSNKKEYLOVDLR	334
		: : : : : : : : : : : : :	
QY	311	KEKKIRGIITG--STMVEHNYYVSAYRILYSDDGOKWTVYREPGEODKIPQGNKDXYHQ	368
		: : : : : : : : : : : : :	
Db	335	FLMLTAIATOGAISRETONGYVVKYSKLEVSTNGEDMMVYRHG--KNHKVFQOANDDATE	392
		: : : : : : : : : : : : :	
QY	369	DVANNEFLPIIARIFRVNVTQOOQTAMKWELLGCOFIPKGPPKLTOPPPPRNSNDLKN	428
		: : : : : : : : : : : : :	
Db	393	VVLINKHAPELTLTRFVRIRPOTHSHGIALTELFGC-----RVTDAPCSNNLMGLMSG	443
		: : : : : : : : : : : : :	
QY	429	TTPAPKIAGRPFKPF-----TOPLOPRSSNEEFPACOTEQTASPDIRNTVTVPNVTKDV--	481
		: : : : : : : : : : : : :	
Db	444	LIADSQISASSTQEYLWSPSAARLVSSRGWFPRIPQAQPGEEWLQVDLGITPKTVKGVII	503
		: : : : : : : : : : : : :	
QY	482	-----ALAAVLPVLYVMVLTLLILVCANH--NRNKKKTEGT--YDLPYWDRA	527
		: : : : : : : : : : : : :	
Db	504	QGARGOSITAVEARAFVRKFVKSYSLNCKDWEYTQDPRTQOPKLFEGNMHHYDTP----	558
		: : : : : : : : : : : : :	
QY	528	GWNKMKQLPALPAVDHEETPVRYSSSEYNHLSPREVTTVLQADSAEY--AQPLVGGIVG	585
		: : : : : : : : : : : : :	
Db	559	-----DIRRPDP-----IPQYRVYPERSPAGIGNRLEVLGCDWTDSKPTVETLGP	606
		: : : : : : : : : : : : :	
QY	586	TLHORSTFKP-----BEGKEAG	602
		: : : : : : : : : : : : :	
Db	607	TVKSEETTTPYPTEEEATECG	627

Search completed: January 21, 2003, 09:53:46
Job time : 29.9573 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:52:40 ; Search time 8.96744 Seconds
(without alignments)
3371.781 Million cell updates/sec

Title: US-10-060-830-3

Perfect score: 3888

Sequence: 1 MPFLFLLLVLLLLLEDAGA.....TOEYSGAGRDGECDFKREIL 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	402.5	10.4	925	1	NRP2_RAT	O35276 rattus norv
2	400.5	10.3	923	1	NRP1_HUMAN	O14786 homo sapien
3	399.5	10.3	931	1	NRP2_MOUSE	O35375 mus musculu
4	395.5	10.2	931	1	NRP2_HUMAN	O60462 homo sapien
5	391.5	10.1	923	1	NRP1_MOUSE	P97333 mus musculu
6	387.5	10.0	922	1	NRP1_RAT	Q96419 rattus norv
7	384.5	9.9	914	1	NRP1_CHICK	P79795 gallus gall
8	383	9.9	928	1	NRP1_XENLA	P28824 xenopus lae
9	362	9.3	2258	1	FA5_PIG	Q96194 sus scrofa
10	361	9.3	2211	1	FA5_BOVIN	Q28107 bos taurus
11	348	9.0	427	1	MFGM_BOVIN	Q95114 bos taurus
12	347	8.9	2224	1	FA5_HUMAN	P12259 homo sapien
13	343	8.8	387	1	MFGM_HUMAN	O08431 homo sapien
14	339	8.7	409	1	MFGM_PIG	P79385 sus scrofa
15	339	8.7	427	1	MFGM_RAT	P70490 rattus norv
16	321	8.3	463	1	MFGM_MOUSE	P12956 mus musculu
17	311	8.0	2133	1	FA8_PIG	P12263 sus scrofa
18	308.5	7.9	2319	1	FA8_MOUSE	O06194 mus musculu
19	308.5	7.9	2351	1	FA8_HUMAN	P00451 homo sapien
20	269.5	6.9	224	1	XLRL_MOUSE	Q92114 mus musculu
21	261.5	6.7	224	1	XLRL_HUMAN	O15537 homo sapien
22	256.5	6.6	280	1	XLRL_FUGRU	Q9W6r5 fuigu brup
23	217	5.6	1331	1	CTA2_HUMAN	Q9uhc6 homo sapien
24	213	5.5	1313	1	HMCT_BOMMO	P98092 bombyx mori
25	208	5.3	1310	1	CTA4_MOUSE	Q99p47 mus musculu
26	207.5	5.3	1284	1	NFX4_DROME	Q94887 drosophila
27	206	5.3	1308	1	CTA4_HUMAN	Q9C0a0 homo sapien
28	200.5	5.2	855	1	DDR2_HUMAN	Q16832 homo sapien
29	198	5.1	277	1	TSG6_HUMAN	P98066 homo sapien
30	198	5.1	707	1	BMP1_XENLA	P98070 xenopus lae
31	193.5	5.0	1022	1	TLD_BRARE	O57460 brachydanio
32	193	5.0	275	1	TSG6_MOUSE	O08859 mus musculu
33	193	5.0	1035	1	ENTK_BOVIN	P98072 bos taurus

34	192.5	5.0	991	1	BMP1_MOUSE	P98063 mus musculu
35	191.5	4.9	986	1	BMP1_HUMAN	P13497 homo sapien
36	189.5	4.9	705	1	C1R_HUMAN	P00736 homo sapien
37	186	4.8	913	1	DDR1_HUMAN	Q08345 h epithelia
38	185.5	4.8	276	1	TSG6_RABIT	P98065 oryctolagus
39	185.5	4.8	688	1	CLS_HUMAN	P09871 homo sapien
40	184.5	4.7	468	1	PCOL_MOUSE	Q81398 mus musculu
41	184.5	4.7	468	1	PCOL_RAT	O08628 rattus norv
42	184.5	4.7	911	1	DDR1_MOUSE	Q03146 mus musculu
43	184	4.7	854	1	DDR2_MOUSE	Q62371 mus musculu
44	182	4.7	616	1	SPAN_STRPU	P98068 strongyloce
45	182	4.7	1288	1	CTA3_HUMAN	Q9Bz76 homo sapien

ALIGNMENTS

RESULT 1

NRP2_RAT

ID NRP2_RAT

AC O35276;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).

GN NRP2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP STRAIN=Sprague-Dawley;

RC MEDLINE=97433085; PubMed=9288754;

RA Kolodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J.,

RA Ginty D.D.;

RT "Neuropilin is a semaphorin III receptor.";

RL Cell 90:753-762(1997).

CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165

CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- TISSUE SPECIFICITY: FOUND IN CERTAIN NEURONAL POPULATIONS OF THE

CC CNS AND IN OTHER NONNEURONAL TISSUES INCLUDING MESENCHYMAL TISSUE

CC LINING IN THE RIBS.

CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.

CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 FAM DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; AF016297; AAC53338.1; -

CC HSP; P12259; 1C2T

DR InterPro; IPR000859; CUB_domain.

DR InterPro; IPR000421; FA58_C.

DR InterPro; IPR000998; MAM_domain.

DR Pfam; PF00431; CUB; 2.

DR Pfam; PF00629; MAM; 1.

DR Pfam; PF00754; F5_F8_type_C; 2.

DR SMART; SM00042; CUB; 2.

DR SMART; SM00231; FA58C; 2.

DR SMART; SM00137; MAM; 1.

DR PROSITE; PS01180; CUB; 2.

DR PROSITE; PS01285; FA58C_1; 2.

DR PROSITE; PS01286; FA58C_2; 2.

DR PROSITE; PS50060; MAM_2; 1.

DR

Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor.
KW SIGNAL 1 22
FT CHAIN 23 925
FT DOMAIN 23 858
FT TRANSMEM 859 883
FT DOMAIN 884 925
FT DOMAIN 128 142
FT DOMAIN 149 267
FT DOMAIN 277 427
FT DOMAIN 434 592
FT DOMAIN 642 802
FT DISULFID 28 55
FT DISULFID 83 105
FT DISULFID 149 175
FT DISULFID 208 230
FT DISULFID 277 427
FT DISULFID 434 592
FT CARBOHYD 152 152
FT CARBOHYD 157 157
FT CARBOHYD 629 629
FT CARBOHYD 833 833
FT CARBOHYD 834 834
SQ SEQUENCE 925 AA; 103896 MW; 3BF62903F644851C CRC64;
Query Match 10.4%; Score 402.5; DB 1; Length 925;
Best Local Similarity 23.8%; Pred. No. 5.6e-20;
Matches 163; Conservative 91; Mismatches 291; Indels 139; Gaps 26;
QY 1 MPEFLLLLVLLLLL--EDAGAAQGDGCGHVLGPESGTLTSLNYPQYPNSTVCEWIR 58
Db 1 MDMPFLTIFLALYFSGHKVRSQDPCGGRNLNSKADGAYITSPGYDPYPSHQNCWVYV 60
QY 59 V-KMGERVRIKFG-DFDIEDSCSFHNYLRYNGIVGSRTEIGKYKGLGLQMN--HSIE 113
Db 61 APEPNQKIVLNPHEIEKHD-CYDFIEDRDGDSADLLGKCG-----NAPPTII 114
QY 114 SKGNEITLLFMSGIHVSGRGFIASVIDKODLITCLDTASNFLPEFSKYPACGLLPF 173
Db 115 SSGSVLYIKFTSDYARQAGFSLRYEIKFTGS-----EDCKNFTSPNGTIESPG----F 165
QY 174 AEISGTIPHGYSPLCNAGVHAGVSVNTLGGQIS-----VVISKGP- 217
Db 166 PE---KYPHNL-DCFTTILAKPRMEILIOFLTLEHDPLOVGEGCKYDLDWDGIPH 221
QY 218 -----YVESLANNVTSVVGHLSTSLTFKTS-----GCY 247
Db 222 VGPLIGKYGCTKTPSKLSSGTGLS---LTFTDMAVAKDGSARYLYLHQPENFOCN 278
QY 248 GTLGMSGVIAADPQTASSVLEWTDHTGOENSWKPKKARLKKPGPPAAAFATDEYQWLQI 307
Db 279 APLGMSGRIANEQISASS-----TFSDGRWTQQSRLHGDNDNGWTPNVDNKNYLVQV 331
QY 308 DLNKKKKTGITG--STMVHNYYVSAIRLYSDCKQKTVVREPVEQDKTFQGNKD 365
Db 332 DUREFTMLTATQAGAISETQKGYVSKYLEVSTNGEDWMVYRHG--KHKHVFQANN 389
QY 366 YHQRVNNFLPPIARFVRNPTQWQKIAMKMLGCGFIPKGRPPKLTOTPPPPRNSD 425
Db 390 ATEVLNKLHPLLRFRIRPQTHLGLARLLEFG-----RVTDAPCSNMUGM 440
QY 426 LKNTTAPPKIAGRAPKF-----TOPLOPRSSNEFFPAOTEOTASPDIRNTVTNPKD 480
Db 441 LSLGIADTQISASSFRELWSPSARLYSSRGWFPNRPQAPQGEWLOVDLGTPTKVG 500
QY 481 V-----AAAVLVPLVWVLTLLILVCAWH-----RNRKKKTEGT--YDLPY 524
Db 501 VTIOGARGGDSITAMEARAFVKFKVSYSLNGKDWEYIQDPTQPKLFEGNMHYDTP-- 558
QY 525 DRAGWNKMKQPLPAKADVHEETPVYSSSEVNHLSPREVTIVLOADSAEY--AQPLVGG 582
Db 559 -----DIRR-----EPVPAQYVRVYPERWSPAGIGMRLEVLGCDWTDKPTVET 603
QY 583 IVGTLHORSTFKP-----BEGKEAG 602

Db 604 LGPTVKSEBETTPYPMDEDATECG 627
RESULT 2
NRPI_HUMAN
ID NRPI_HUMAN STANDARD; PRT; 923 AA.
AC O14786; O60461;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuropilin-1 precursor (Vascular endothelial cell growth factor 165 receptor).
GN NRPI OR NRP OR VEGF165R.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM).
MEDLINE=97433084; PubMed=9288753;
He Z., Tessier-Lavigne M.;
RA "Neuropilin is a receptor for the axonal chemorepellent semaphorin III";
RT Cell 90:739-751(1997).
RL [2]
SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM), AND SEQUENCE OF 22-39.
RC TISSUE=Breast;
RX MEDLINE=98188099; PubMed=9529250;
RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745(1998).
[3]
SEQUENCE FROM N.A. (SOLUBLE/SNRP1 ISOFORM), AND SEQUENCE OF 22-31.
RC TISSUE=Prostatic adenocarcinoma;
RX MEDLINE=20183929; PubMed=10688880;
RA Gagnon M.L., Bielenberg D.R., Gechtman Z., Miao H.-Q., Takashima S., Soker S., Klagsbrun M.;
RT "Identification of a natural soluble neuropilin-1 that binds vascular endothelial growth factor: In vivo expression and antitumor activity";
RT Proc. Natl. Acad. Sci. U.S.A. 97:2573-2578(2000).
[4]
CHARACTERIZATION
MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Foltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid form of vascular endothelial growth factor (VEGF) and of placenta growth factor-2, but only neuropilin-2 functions as a receptor for the 145-amino acid form of VEGF";
RL J. Biol. Chem. 275:18040-18045(2000).
[5]
FUNCTION: THE MEMBRANE-BOUND ISOFORM IS A RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGF, THE VEGF-165 ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS. IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS.
[6]
FUNCTION: THE SOLUBLE/SNRP1 ISOFORM BINDS VEGF-165 AND APPEARS TO INHIBIT ITS BINDING TO CELLS. IT MAY ALSO INDUCE APOPTOSIS BY SEQUESTERING VEGF-165. MAY BIND AS WELL VARIOUS MEMBERS OF THE SEMAPHORIN FAMILY. ITS EXPRESSION HAS AN AVERSE EFFECT ON BLOOD VESSEL NUMBER AND INTEGRITY.
[7]
SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. THE SOLUBLE/SNRP1 ISOFORM IS SECRETED.
[8]
ALTERNATIVE PRODUCTS: 2 ISOFORMS; MEMBRANE-BOUND (SHOWN HERE) AND SOLUBLE/SNRP1; ARE PRODUCED BY ALTERNATIVE SPLICING.
[9]
TISSUE SPECIFICITY: THE EXPRESSION OF MEMBRANE-BOUND AND SOLUBLE/SNRP1 ISOFORMS DOES NOT SEEM TO OVERLAP. THE MEMBRANE-BOUND ISOFORM IS EXPRESSED BY THE BLOOD VESSELS OF DIFFERENT

QY	6	LLLLLLLLL	DAGA	QCGCHTV	LGPS	GTILTS	INYPQT	PNSTVC	EWETR	V-KMGR	64							
											:							
Db	7	LLCAV	ALVLA	PAGAF	RNDK	CGDTI	KIES	PGYLT	SPGYH	SPKSC	EWLIQADPDYQ	66						
											:							
QY	65	VRKFG	-DPD	IEDSD	CHFN	LYRI	ANGI	GVSR	TEIG	KYCG	LG	GLQ	MNHS	TES	KGNET	TLF	123	
																	:	
Db	67	IMIN	PHF	DEL	D	R	D	-CKY	D	Y	E	V	F	V	E	G	E	123
																	:	
QY	124	MSGI	HVSG	RFL	AS	Y	VIDK	-----	QDL	T	-----	-----	-----	-----	-----	-----	148	
																	:	
Db	124	VSDY	ETH	G	A	G	F	S	Y	E	I	R	G	P	C	S	Q	183
																	:	
QY	149	--	CLD	T	A	S	N	L	E	P	E	F	S	K	Y	P	A	202
																	:	
Db	184	EI	L	E	F	E	S	F	D	E	P	D	S	N	P	--	P	237
																	:	
QY	203	T	L	G	O	I	S	V	W	I	S	K	G	I	P	Y	E	257
																	:	
Db	238	S	S	G	I	L	S	W	F	-----	Y	D	S	A	I	A	K	286
																	:	
QY	258	A	D	P	Q	I	T	A	S	S	V	E	L	T	D	T	G	317
																	:	
Db	287	H	S	D	I	T	A	S	S	-----	Q	Y	S	T	N	S	A	338
																	:	
QY	318	I	I	T	G	--	S	T	W	E	N	H	Y	S	A	R	I	375
																	:	
Db	339	V	G	T	O	G	A	I	S	K	E	T	K	K	Y	Y	K	396
																	:	
QY	376	P	P	I	A	R	I	R	N	P	T	O	W	O	K	I	A	418
																	:	
Db	397	K	L	T	R	V	R	I	K	A	T	W	E	G	I	S	M	430
																	:	

RESULT 3

NRP2_MOUSE

NRP2_MOUSE

STANDARD;

PRT: 931 AA

O35375; O35373; O35374; O35376; O35377; O35378;

16-OCT-2001 (Rel. 40, Created)

16-OCT-2001 (Rel. 40, Last sequence update)

15-JUN-2002 (Rel. 41, Last annotation update)

Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).

DE NRP2.

GN

OS Mus musculus. (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxId=10090;

RN [1]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RC STRAIN=BALB/c;

RX MEDLINE=97470888; PubMed=9331348;

RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;

RT

[2]
 RN SEQUENCE FROM N.A. (ISOFORM A22).
 RC TISSUE-Breast;
 RA MEDLINE=9818099; PubMed=9529250;
 RX Soker S., Takashima S., Miao H.-O., Neufeld G., Klagabrun M.;
 "Neuropilin-1 is expressed by endothelial and tumor cells as an
 RT isoform-specific receptor for vascular endothelial growth factor.";
 RL Cell 92:735-745(1998).
 RP CHARACTERIZATION.
 RA MEDLINE=20309748; PubMed=10748121;
 RX Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
 "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
 RT form of vascular endothelial growth factor (VEGF) and of placenta
 FT growth factor-2, but only neuropilin-2 functions as a receptor for
 the 145-amino acid form of VEGF";
 RL J. Biol. Chem. 275:18040-18045(2000).
 CC -1- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
 CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
 CC -1- SUBUNIT: NEUROFILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
 CC NEUROFILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN 3C RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: A22 (SHOWN HERE), A0 AND A17;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF022859; AAC51788.1; -;
 DR EMBL; AF022860; AAC51789.1; -;
 DR EMBL; AF016098; AAC12922.1; -;
 DR HSP; P12259; 1C2T.
 DR Genew; HGNC:8005; NRP2.
 DR MIM; 602070; -;
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR000998; MAM_domain.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00629; MAM; 1.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; SM00137; MAM; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00600; MAM_2; 1.
 KW Transmembrane; Glycoprotein; Neuron; Signal; Repeat; Receptor;
 KW Alternative splicing.
 FT SIGNAL 1 20 OR 22 (POTENTIAL).
 FT CHAIN 21 931 NEUROFILIN-2.
 FT DOMAIN 21 864 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 865 889 POTENTIAL.
 FT DOMAIN 890 931 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 28 142 CUB 1.
 FT DOMAIN 149 267 CUB 2.
 FT DOMAIN 277 427 F5/8 TYPE C 1.
 FT DOMAIN 434 592 F5/8 TYPE C 2.
 FT DOMAIN 642 802 MAM.
 FT DOMAIN 671 674 POLY-SER.
 FT DISULFID 28 55 BY SIMILARITY.
 FT DISULFID 83 105 BY SIMILARITY.
 FT DISULFID 149 175 BY SIMILARITY.
 FT DISULFID 208 230 BY SIMILARITY.

FT DISULFID 277 427 BY SIMILARITY.
 FT DISULFID 434 592 BY SIMILARITY.
 FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 839 839 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 809 813 MISSING (IN ISOFORM A17).
 FT VARSPPLIC 809 830 MISSING (IN ISOFORM A0).
 FT CONFLICT 602 602 E -> K (IN REF. 1).
 SQ SEQUENCE 931 AA; 104830 MW; 270CBAE69A0A797C CRC64;
 Query Match 10.2%; Score 395.5; DB 1; Length 931;
 Best Local Similarity 23.6%; Pred. No. 1.7e-19;
 Matches 161; Conservative 92; Mismatches 299; Indels 133; Gaps 25;
 QY 1 MPFL--LLVLLVLLLEDAAGQGGGCVLGPESGLTSTINYPQYPNSTVCEWEIR 58
 DB 1 MDMFLLTWVFLALYSRHRVGRQPPPCGRLNSKDAVITSPGYPDYPSHQNCWIVY 60
 QY 59 V-KMGERVRIKRG-DFDIEDSDSCHFNLYRIYNGVSRTEIGKYCGGLQMN---HSIE 113
 DB 61 APEPNQKIVLNFPHFEIEKHD-CRYDFEIRDGDSADLLGKHG-----NIAPPTII 114
 QY 114 SKGNETLLFMGSIHVSGRGLASYSVIDKQDLTCLDTASNFLEPEFSKYCPAGCLLPF 173
 DB 115 SSGSMLYIKFTSDYAROGAGFSIRYEIFKGS-----EDCSKNFTSPNGTIESPG- 165
 QY 174 ABISGTIPHGYSRSPCLMCAHVAGVWSNLTGGQIS-----VVISKGIP- 217
 DB 166 PE--KYPHNL-DCFTILAKPKMEILLOFLFDLHDPLOVGECDKYDWDLDWIGPH 221
 QY 218 -----YVESLANNVTSVGHLSLPT-----FTSGCYGTL 250
 DB 222 VGPLIGKYCGTKTPSLRSLSTGILSTFHTDMAVADGFSARYYLHVQEPLENFQCNVPL 281
 QY 251 GMESGVADPQITASSVLEWTDHTGQENSWKPKARLKKPGPPWAAFAFATDEYQWLIDLN 310
 DB 282 GMESGRIANEQISASS-----TSDGRWTQQSKRLHGDGNGWTPNLDNSKEYLQVDLR 334
 QY 311 KEKKTIGITTG--STMVEHNYVSAYRILYSDDGKMTVYREPGVEQDKIFQGNKDYHQ 368
 DB 335 FLTMLTAIATQGAISRETONGYVSKYLEVSTNGEDWVYRHG--KNHKVQANNDAT 392
 QY 369 DVRNNEFLPIIARFIRVNPTQOQKAMKELLGCOFIPKGRPPKLTQPPPPRNSNDLKN 428
 DB 393 VVLNKLHAPLLTRFVIRPQTWHSGIALRLFLFC-----RVTDAPCSNMLGMLSG 443
 QY 429 TTAPPKIAKGRAPK-----TOPQPRSNRNEPPAQTEQTASPDINTTVPNTKDV-- 481
 DB 444 LIADSOISASSTOEYLMWSPSAARLVSSRSGWFPRIPOAQPGBEWLQVDLGTPTVKGVII 503
 QY 482 -----ALAAVLVPVLMVLTLLILVCAWHW-----RNRKKKTEGT--YDLPYWDRA 527
 DB 504 QGARGGDSITAVEARAFVRFKVSYSLNGKDWEYIODPRTQPKLEGNHMYDTP----- 558
 QY 528 GWMKGMKQFLPAKAVDHEETPVYSSSENVHLSPREVTTVLQADSAY--AOLVGGIYG 585
 DB 559 ----DIRREDP-----IPAQYVRVYPERWSPAGIGMRLEVLCGDWTDTSKPTVETLGP 606
 QY 586 TLHQRSTFKP-----EEGKEAG 602
 DB 607 TVKSEETTTPTPEEEATECG 627
 RESULT 5
 NRPL_MOUSE
 ID NRPL_MOUSE STANDARD; PRT; 923 AA.
 AC P97333;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 GN Neuropilin-1 precursor (A5 protein).
 GN NRPI OR NRP.

OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALE/c; TISSUE=Embryonic brain;
 RX MEDLINE=96353149; PubMed=8748368;
 RA Kawakami A., Kitsuawa T., Takagi S., Fujisawa H.;
 RT "Developmentally regulated expression of a cell surface protein,
 RL J. Neurobiol. 29:1-17(1996).
 CC -!- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
 CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT
 CC BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGF. THE VEGF-165
 CC ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN
 CC INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS.
 CC IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: NERVOUS SYSTEM.
 CC -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D50086; BAA08789.1; -.
 DR HSP; P12259; ICYT.
 DR MGD; MGI:106206; Nrp.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR000998; MAM_domain.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00629; MAM; 1.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; SM01137; MAM; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00740; MAM_1; 1.
 DR PROSITE; PS00060; MAM_2; 1.
 KW Transmembrane; Glycoprotein;
 FT SIGNAL 1 21
 FT CHAIN 22 923
 FT DOMAIN 22 856
 FT TRANSMEM 857 879
 FT DOMAIN 880 923
 FT DOMAIN 27 141
 FT DOMAIN 147 265
 FT DOMAIN 275 424
 FT DOMAIN 431 583
 FT DOMAIN 645 811
 FT DISULFID 27 54
 FT DISULFID 82 104
 FT DISULFID 147 173
 FT DISULFID 206 228
 FT DISULFID 275 424
 FT DISULFID 431 583
 FT CARBOHYD 150 150
 FT CARBOHYD 261 261
 FT CARBOHYD 300 300
 FT CARBOHYD 522 522

FT CARBOHYD 842 842 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 923 AA; 103020 MW; 0644B8A170796808 CRC64;
 Query Match 10.1%; Score 391.5; DB 1; Length 923;
 Best Local Similarity 27.3%; Pred. No. 3.2e-19;
 Matches 129; Conservative 60; Mismatches 176; Indels 107; Gaps 18;
 QY 6 LLLVLLVLLLEDAGAQGGGHTVLPESGTLTSLTINYPQYPSNTVCEWEIRV-KMGER 64
 DB 7 LLCTATLALALAGAFSDKGGTKIKIENPGVLTSPGYSPHSPKCEWLIIQAEPEYQR 66
 QY 65 VRIKFG-DFDIEDSDSCHFNLYRNYNGVSRTEIKYCKGLGLQNMHS-ISKGNIEILL 122
 DB 67 IINFNPHFDLEDRD-CKYDVEVIDGENEGRLMGKFCG---KIAPSPVSSGPFLEIK 122
 QY 123 FMSGIHVSGRGLASYSVIDKODLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTIPH 182
 DB 123 FVSDYETHGAGFSRYEIFKRGP-----ECQNYTAP-----TGVIKS 160
 QY 183 -GYRDSGLPCM-----AGVHAGVV-----SNTLGGQI-----SVVISKIP-- 217
 DB 161 PGFPEKYPNCLCTYIIFAPKMSIILEFESFDLEQDSNPPGCMFCYDRLEIMDGFPEV 220
 QY 218 -----YYESSLANNVTSVVGHLSTSLT-----FRTSGCYG 248
 DB 221 GPHIGRCGQKTPGRISSSGVLSMVFYDTSIAIAKEGFSANYSVLQSSISEDFK---CME 277
 QY 249 TLGMESGVIAQPQTASSVLEWTDHTGOENSKPKKARLKPFPWAAAFATDEYQWLQID 308
 DB 278 ALGMESGEIHSQDITASS-----QYGTNWSVERSLNYPENGWTPGEDSYKEWIQVD 329
 QY 309 LNKEKKITGITTG--STWVEHNYVVSAYRILYSDGQKWTYRPPGVQDKIFQGNKDY 366
 DB 330 LGLLRFVTAVGTAISRETKKKYVYTRYVDISSNGEDWISLKEG--NKALIFQGNTP 387
 QY 367 HDVRRNFPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPRKPTQPP 418
 DB 388 TDVVLGVFSKPLITRFVRIKPVSWETGISMRFEVYGC-----KTDYP 430
 RESULT 6
 NRPI_RAT ID NRPI_RAT STANDARD; PRT; 922 AA.
 AC Q90WJ9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neuropilin-1 precursor (Vascular endothelial cell growth factor 165
 DE receptor).
 GN NRPI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley;
 RX MEDLINE=97433085; PubMed=9288754;
 RA Kolodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J.,
 RA Ginty D.D.;
 RT "Neuropilin is a semaphorin III receptor."
 RL Cell 90:753-762(1997).
 CC -!- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
 CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT
 CC BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGF. THE VEGF-165
 CC ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN
 CC INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS.
 CC IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: FOUND IN THE EMBRYONIC NERVOUS SYSTEM.
 CC -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.

DR PRINTS; PRO0020; MAMDOMAIN.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; SM00137; MAM; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00740; MAM_1; 1.
 DR PROSITE; PS00060; MAM_2; 1.
 KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
 KW Cell adhesion.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 914 NEUROFILIN-1.
 FT DOMAIN 20 847 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 848 870 POTENTIAL.
 FT DOMAIN 871 914 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 25 139 CUB 1.
 FT DOMAIN 145 263 CUB 2.
 FT DOMAIN 273 422 F5/8 TYPE C 1.
 FT DOMAIN 429 581 F5/8 TYPE C 2.
 FT DOMAIN 636 801 MAM.
 FT DISULFID 25 52 PROBABLE.
 FT DISULFID 80 102 PROBABLE.
 FT DISULFID 145 171 PROBABLE.
 FT DISULFID 204 226 PROBABLE.
 FT DISULFID 273 422 BY SIMILARITY.
 FT DISULFID 429 581 BY SIMILARITY.
 SQ SEQUENCE 914 AA; 102480 MW; DD2EE6D6F0CB68C CRC64;

Query Match 9.9%; Score 384.5; DB 1; Length 914;
 Best Local Similarity 26.1%; Pred. No. 9.6e-19;
 Matches 131; Conservative 72; Mismatches 200; Indels 99; Gaps 19;
 QY 20 AAGGCGGHTVLPSPSGILTSINYQTPNNTVCWEIRV-KMGSRVIRKFG-DFDIIDS 77
 Db 19 ALRSKCGTITKILSPGLTSPGYQSPYHPSQKCEWLQAPEYQIRIMINPNPHDLEDR 78
 QY 78 DSCFNLYRIYIGVSRTEICKYGLGLQNMHSIESKNEITLLFMSGIHVSIGRGFLAS 137
 Db 79 D-CKDYVEVIDGNAEGLGKYC--GKIAPPPLVSSGPLYFIKFSVDYETHGAGFSIR 135
 QY 138 YSVIDK-----QDLITC-----LDTASNFLPE 160
 Db 136 YEVFKRGPECSRNTSSSGMIKSPGFPEKYPNSLCTYIIFAPKMSIILEFSEFLEPD 195
 QY 161 FSKYCPACGLLPFAISGTHPGYRDSPP-----LCMAGVHAGVSVNTLGGQISVVISKGI 216
 Db 196 SN--TPGGAFCRYDRLE--INDGPDVGVPHIGRYCGQNPGVRSEST--GILSMVF----- 245
 QY 217 PYESSLANNVTSVVGHLSLSTLFTKTSCTGLGMSGVIAADPOITASSVLEWTDHTGQ 276
 Db 246 -YDTSIAIAKEGFSANYSVSQSSVS--EDFQCEPLNGSGEIHSDQITVSS-----Q 294
 QY 277 ENS-WKPKKARKLKPGPPAAFAFADYQWLQIDLNKKKITGIITG--STMVHNHYVS 333
 Db 295 YSAIWSSERSRLNYPENCWTGCDSDVREWIQVGLLLRFVSSIGTQGAISKETKEYLYK 354
 QY 334 AYRILYSDGQKWTYRPEGVQEDKIFQGNKDYHODVRNNLPPIIARFIRVNPQWQOK 393
 Db 355 TRDVDSNGEDWITLKEG--NKPVVYFQGNSTPDVYRPPKPVLTFRVRIKPVSWENG 412
 QY 394 IAMKMLGCGQIPKGRPKLTQPP-----PPRNSDLKNTAPPKTAGKRAP 441
 Db 413 VSLREVYGC-----KITDYPCSMGLMGVSLIPD-----SQITASTQVDRNWIP 457
 QY 442 KFTQPLQPRSSNEFFAQEQIT 463
 Db 458 ENARLITRSGLWALPTTHPYT 479

RESULT 8
 NRPI_XENLA
 ID NRPI_XENLA STANDARD; PRT; 928 AA.

AC P28824;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neuropilin-1 precursor (A5 protein) (A5 antigen).
 OS Xenopus laevis (African Clawed frog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Eupidae;
 CC Xenopodinae; Xenopus.
 CC NCBI_TaxID=83355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA MEDLINE=91337458; PubMed=1908252;
 RA Takagi S., Hirata T., Agata K., Mochii M., Eguchi G., Fujisawa H.;
 RT "The A5 antigen, a candidate for the neuronal recognition molecule,
 has homologies to complement components and coagulation factors.";
 RL Neuron 7:295-307(1991).
 CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
 CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
 CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOREGULATION OF THE NERVOUS
 CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
 CC SIMILARITY). PRESUMED TO BE INVOLVED IN THE NEURONAL RECOGNITION
 CC BETWEEN THE OPTIC NERVE FIBERS AND THE VISUAL CENTERS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: RETINAL GANGLION CELLS AND VISUAL CENTER
 CC NEURONS.
 CC -1- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10467; BAA01260.1; -
 DR HSPSP; P12259; LC2T.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR000998; MAM_Domain.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00629; MAM; 1.
 DR Pfam; PF00754; F5_P8_typeC; 2.
 DR PRINTS; PR00020; MAMDOMAIN.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; SM00137; MAM; 1.
 DR PROSITE; PS00740; MAM_1; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 DR PROSITE; PS00060; MAM_2; 1.
 KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
 KW Antigen.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 928 NEUROFILIN-1.
 FT DOMAIN 22 860 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 861 883 POTENTIAL.
 FT DOMAIN 884 928 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 27 141 CUB 1.
 FT DOMAIN 147 265 CUB 2.
 FT DOMAIN 275 424 F5/8 TYPE C 1.
 FT DOMAIN 431 584 F5/8 TYPE C 2.
 FT DOMAIN 646 812 MAM.
 FT DISULFID 27 54 PROBABLE.
 FT DISULFID 82 104 PROBABLE.
 FT DISULFID 147 173 PROBABLE.
 FT DISULFID 206 228 PROBABLE.

```

FT DISULFID 275 424 BY SIMILARITY.
FT DISULFID 431 584 BY SIMILARITY.
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 928 AA; 103416 MW; AF6B323B0A4C789D CRC64;

Query Match 9.9%; Score 383; DB 1; Length 928;
Best Local Similarity 25.8%; Pred. No. 1.2e-18;
Matches 138; Conservative 70; Mismatches 197; Indels 130; Gaps 22;

Qy 5 LLLLLV-----LLLLLDAGAQGDGCGHTVLGPESGILTSINYQTPVNPSTVCWEIRV 59
Db 1 MLLRLSCCCWLLSLRSWSARNDKGDTKITSPSYLTSGAGYPHSPPSQRCWEILQA 60

Qy 60 -KMGVERIKFG-DFDIEDSDSCHFNLYRIYNGIGVSTETGKYGLGLQNMHS-IESKG 116
Db 61 PEHYQIRIMINPHFDLEDE-CKDYIYVEVDGDNANGQLLGKYGKCG---KIAPSLVSTG 116

Qy 117 NEITLLFMGSHVSGRGFLASVYIDKQDLITCLDTASNFLEPEFSKYCPACGCLLPFAEI 176
Db 117 PSIFIRFVSDYETPGAGFSIRYEVFK-----TCPE-----CSRFTSS 154

Qy 177 SGTI-----PHGYRDS---SPLCMAGVHAGV-----SNTLGGQIS-----VVIS 213
Db 155 NGVIKSPKYPEKYPNALECTYIIFAPKMQEIVLEFESFELEADSNAPGGQTCRYDWLGTW 214

Qy 214 KGIP-----YYESSLANNVTSVGHLSLTFKTSFG-----245
Db 215 DGFPGVGHIGRYCQNPGRVRSPTGILSMIFHDSIAIEKGFANFSVQSNVDSDEDFQ 274

Qy 246 CYGTLMGMSGVIADPQITASS--VLEWTDHRCQ-----ENSWKPKKARLKKPGPPWAAFA 299
Db 275 CKEALMGESGEIHFQIISVSQISYNNWSAERSRLNVENGWTPGEDTVK-----323

Qy 300 DEYQWLQDLNKEKITGIITG--STMVEHNYYSATRIYSDGQKWTYVREPGVQD 357
Db 324 ---EWIQVDLENLRFVSGIGTQGAISKETKKYFKVSKYKVDISSNGEDWITLKDG--NKH 378

Qy 358 KIFQGNKDYHODVRNRPPLPIIARIVNPTQWQOKIAMKMLLGCQIPKGRPPKLTQP 417
Db 379 LVFTGNTDQTVVRPFPKPVITREVRURPTWENGISIRFELIYC-----KIDY 429

Qy 418 PPPRNSDL-----KNTPAPPKIAKGRAPKFTQPLQPRS-----SNEFFAQTE 460
Db 430 PCSRLMGVSGLISDSQITASSQVDNRNVPELARLVTSRSGWALPPSNTHPYTKE 484

RESULT 9
FA5_PIG ID FA5_PIG STANDARD; PRT; 2258 AA.
AC O9GLPL;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
[1]
NCBI_TaxID=9823;
RN SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C
RP DOMAINS.
RC TISSUE=Liver;
RX MEDLINE=21121490; PubMed=11229814;
RA Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,
RA Kim H.K.W.;
RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional
RT protein modeling of membrane binding sites and comparative anatomy of
RT domains.";

```

```

RL Cell. Mol. Life Sci. 58:148-159(2001).
CC -|- FUNCTION: Coagulation factor V is a cofactor that participates
CC with factor Xa to activate prothrombin to thrombin.
CC -|- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC chain, noncovalently bound. The interaction between the two chains
CC is calcium-dependent.
CC -|- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1
CC and C2 may be involved in membrane binding.
CC -|- PTM: Thrombin activates factor V proteolytically to the active
CC cofactor, factor Va (formation of a heavy chain at the N-
CC terminus and a light chain at the C-terminus).
CC -|- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -|- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -|- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AF191308; AAC28381.1; .
CC HSP; P12259; ICZT.
CC InterPro; IPR001117; Cu-oxidase.
CC InterPro; IPR000421; FA58_C.
CC Pfam; PF00394; Cu-oxidase; 3.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00231; FAS8C; 2.
CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
CC PROSITE; PS01285; FAS8C_1; 2.
CC PROSITE; PS01286; FAS8C_2; 2.
CC Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
CC Repeat.
CC SIGNAL 1 22
CC CHAIN 23 2258
CC CHAIN 23 737
CC PEPTIDE 738 1611
CC
CC CHAIN 1612 2258
CC DOMAIN 30 329
CC DOMAIN 30 193
CC DOMAIN 203 329
CC DOMAIN 348 683
CC DOMAIN 348 525
CC DOMAIN 535 683
CC DOMAIN 691 1611
CC DOMAIN 1168 1539
CC
CC REPEAT 1168 1176
CC REPEAT 1177 1185
CC REPEAT 1186 1194
CC REPEAT 1195 1203
CC REPEAT 1204 1212
CC REPEAT 1213 1221
CC REPEAT 1222 1230
CC REPEAT 1231 1239
CC REPEAT 1240 1248
CC REPEAT 1249 1257
CC REPEAT 1258 1266
CC REPEAT 1267 1275
CC REPEAT 1276 1284
CC REPEAT 1285 1293
CC REPEAT 1294 1302
CC REPEAT 1303 1311
CC REPEAT 1312 1320
CC REPEAT 1321 1329
CC REPEAT 1330 1338
CC REPEAT 1339 1347
CC REPEAT 1348 1356
CC REPEAT 1357 1365
CC
CC POTENTIAL.
CC COAGULATION FACTOR V.
CC HEAVY CHAIN (BY SIMILARITY).
CC ACTIVATION PEPTIDE (CONNECTING REGION)
CC (BY SIMILARITY).
CC LIGHT CHAIN (BY SIMILARITY).
CC F5/8 TYPE A 1.
CC PLASTOCYANIN-LIKE 1.
CC PLASTOCYANIN-LIKE 2.
CC F5/8 TYPE A 2.
CC PLASTOCYANIN-LIKE 3.
CC PLASTOCYANIN-LIKE 4.
CC B.
CC 41 X 9 AA APPROXIMATE TANDEM REPEATS OF
CC T-L-S-P-D-L-[GS]-[HQ]-T.
CC 1.
CC 2.
CC 3.
CC 4.
CC 5.
CC 6.
CC 7.
CC 8.
CC 9.
CC 10.
CC 11.
CC 12.
CC 13.
CC 14.
CC 15.
CC 16.
CC 17.
CC 18.
CC 19.
CC 20.
CC 21.
CC 22.

```


FT	MOD_RES	1537	1537	SULFATION (POTENTIAL).	
FT	MOD_RES	1541	1541	SULFATION (POTENTIAL).	
FT	CARBOHYD	225	225	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	297	297	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	382	382	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	460	460	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	553	553	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	587	587	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	745	745	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	756	756	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	774	774	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	780	780	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	902	902	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	952	952	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	964	964	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1044	1044	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1053	1053	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1062	1062	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1071	1071	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1078	1078	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1094	1094	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1451	1451	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1490	1490	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1550	1550	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1690	1690	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1839	1839	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	1997	1997	N-LINKED (GLCNAC.	(POTENTIAL).
FT	CARBOHYD	2196	2196	N-LINKED (GLCNAC.	(POTENTIAL).
FT	VARIANT	587	592	NTPLPA -> T (IN VARIANT 2).	
SEQ	SEQUENCE	2211	AA; 248981	MW; CBBF90B738667C45	CRC64;
Query Match					
Best Local Similarity		9.3%; Score 361; DB 1; Length 2211;			
Matches 79; Conservative 17; Mismatches 60; Indels 10; Gaps					
Qy	242	KTSCGCTGLMGESVIAADPOITASSVLE--WDHTQGSNRPKPKARLKKPG--PPWAAF	297		
Db	2049	EVNGCSTPLMGESKIKENQITASSPKKSW-----GNYWEPFLARLNAQGRVNAQAK	2101		
Qy	298	ATDEYONLQIDLNKEKITGIITGSTVMEHNYVVSAYRTLYSDDCGKVTVVRPGVEOD	357		
Db	2103	ANNQNWQLDILKKIKITAIVTQCGKLSSESMVYKSYTIHYSDDQGTDKWPREKSSMYD	2161		
Qy	358	KIFQGNKDYHQDVRRNLPPIIAIRFRVNPQTQOQKIAMKMLLGC	403		
Db	2163	KIFGNVNRGHVKNFFNPPIISRFIRIIPKTNQSIARLRLFGC	2208		
RESULT 11					
MFGM_BOVIN					
AC	Q95114; Q27959; STANDARD; PRT; 427	AA.			
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)				
DE	Mcp57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein				
DE	SP47) (BP47) (Components 15/16).				
OS	MFG-E8.				
GN	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID=9913;				
RN	[1]				
RP	SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.				
RC	STRAIN-Holstein; TISSUE-Mammary gland;				
RX	MEDLINE=97008954; PubMed=8856064;				
RA	Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,				
RA	Petersen R.E.;				
RT	"Characterization of glycoprotein PAS-6/7 from membranes of bovine				
RT	milk fat globules.";				

RL Eur. J. Biochem. 240:628-636(1996).
 RN [2]
 RP SEQUENCE OF 18-427 FROM N.A.
 RC TISSUE-Mammary gland;
 RX MEDLINE-96125736; PubMed-8541316;
 RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
 RA Matsuda T.;
 RT "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
 RT monoclonal antibodies raised against bovine milk fat globule
 RT membrane";
 RL Biochim. Biophys. Acta 1245:385-391(1995).
 RN [3]
 RP SEQUENCE OF 19-427 FROM N.A.
 RC TISSUE-Testis;
 RA Ensslin M.A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
 RC TISSUE-Milk;
 RX MEDLINE-93250576; PubMed-8485470;
 RA Mather I.H., Banghart L.R., Lane W.S.;
 RT "The major fat-globule membrane proteins, bovine components 15/16 and
 RT guinea-pig GP 55, are homologous to MGP-E8, a murine glycoprotein
 RT containing epidermal growth factor-like and factor V/VIII-like
 RT sequences";
 RL Biochem. Mol. Biol. Int. 29:545-554(1993).
 CC -!- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
 CC MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
 CC BINDING PROTEIN
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
 CC LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN.
 CC -!- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
 CC -!- PTM: THE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH
 CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.
 CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X91895; CAA62997.1; -;
 DR EMBL; S80643; AAB35894.2; -;
 DR EMBL; Y11719; CAA72406.1; -;
 DR HSRP; P00740; LIXA.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000421; FA58_C.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00754; F5_F8_typeC; 2.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01285; FA58C_1; 2.
 DR PROSITE; PS01286; FA58C_2; 2.
 KW Signal; Glycoprotein; Milk; Repeat; EGF-like domain;
 KW Alternative splicing.
 FT SIGNAL 1 18
 FT CHAIN 19 427 LACTADHERIN.
 FT DOMAIN 20 59 EGF-LIKE 1.
 FT DOMAIN 62 106 EGF-LIKE 2.
 FT DOMAIN 109 265 F5/8 TYPE C 1.
 FT DOMAIN 270 427 F5/8 TYPE C 2.
 FT SITE 85 87 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 24 35 BY SIMILARITY.
 FT DISULFID 29 47 BY SIMILARITY.
 FT DISULFID 49 58 BY SIMILARITY.
 FT DISULFID 66 77 BY SIMILARITY.

FT DISULFID 71 94 BY SIMILARITY.
 FT DISULFID 96 105 BY SIMILARITY.
 FT DISULFID 109 265
 FT DISULFID 252 256
 FT DISULFID 270 427
 FT CARBOHYD 27 27 O-LINKED (FUC. . .) (IN PAS-6).
 FT CARBOHYD 34 34 O-LINKED (FUC. . .) (IN PAS-7).
 FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (HYBRID) (IN PAS-6
 FT CARBOHYD 227 227 AND PAS-7).
 FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (HIGH MANNOSE) (IN
 FT VARSPLIC 169 221 PAS-6).
 FT CONFLICT 19 19 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 28 28 A -> F (IN REF. 1).
 SQ SEQUENCE 427 AA; 47411 MW; 4CBEE3A1DC4EB24 CRC64;
 Query Match 9.08; Score 348; DB 1; Length 427;
 Best Local Similarity 42.08; Pred. No. 1.1e-16;
 Matches 73; Conservative 24; Mismatches 69; Indels 8; Gaps 4;
 QY 234 LSTSLFTFKTSGCYGTGLGMSGVIADPOITASSVLE-WTDHTG-QENSWKPKKARKKPG 291
 DB 258 LRFELLCGELNGCTEPLGKNDTPNKQITASSYKTW----GLSAFSEFPYARLDNQG 313
 QY 292 P--PWAAPATDEYQWLQIDLNKKKITGITSTGMVHNHYVVSAYRILYSDDGKQWVY 349
 DB 314 KFNATAQTNSASEWLQIDLGSKRVGTGIIQTGARDFGHQVAAAYRVAYGDDGVWTVEY 373
 QY 350 REPVEQDKIFQGNKDYHODVNNFPLPIIARFIRVNTQWQOKTAMKMLGCG 403
 DB 374 KDPGASESKIPPGNNMNNSHKKNIFETPFQARFVRIQPVAMHNRTILRVELLGC 427
 RESULT 12
 FA5_HUMAN STANDARD; PRT; 2224 AA.
 ID FA5_HUMAN
 AC P12259; Q14285;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Coagulation factor V precursor (Activated protein C cofactor).
 GN F5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92232668; PubMed=1567832;
 RA Cripe L.D., Moore K.D., Kane W.H.;
 RT "Structure of the gene for human coagulation factor V.";
 RL Biochemistry 31:3777-3785(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87260886; PubMed=3110773;
 RA Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
 RA Hewick R.M., Kaufman R.J., Mann K.G.;
 RT "Complete cDNA and derived amino acid sequence of human factor V.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
 RN [3]
 RP SEQUENCE OF 1-1600 FROM N.A.
 RX MEDLINE-88107560; PubMed=2827731;
 RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
 RT "Cloning of cDNAs coding for the heavy chain region and connecting
 RT region of human factor V, a blood coagulation factor with four types
 RT of internal repeats.";
 RL Biochemistry 26:6508-6514(1987).
 RN [4]
 RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
 RX MEDLINE-86313665; PubMed=3092220;
 RA Kane W.H., Davie E.W.;
 RT "Cloning of a cDNA coding for human factor V, a blood coagulation
 RT factor homologous to factor VIII and ceruloplasmin.";


```
FT DOMAIN 1185 1501 35 X 9 AA APPROXIMATE TANDEM REPEATS OF
FT REPEAT 1185 1193 [TNP]-L-S-P-D-L-S-Q-T.
FT REPEAT 1193 1201 2-1.
FT REPEAT 1201 1202 2-2.
FT REPEAT 1202 1211 2-3.
FT REPEAT 1211 1220 2-4.
FT REPEAT 1220 1229 2-5.
FT REPEAT 1229 1238 2-6.
FT REPEAT 1238 1247 2-7.
FT REPEAT 1247 1256 2-8.
FT REPEAT 1256 1265 2-9.
FT REPEAT 1265 1274 2-10.
FT REPEAT 1274 1283 2-11.
FT REPEAT 1283 1292 2-12.
FT REPEAT 1292 1301 2-13.
FT REPEAT 1301 1310 2-14.

Query Match 8.9%; Score 347; DB 1; Length 2224;
Best Local Similarity 45.8%; Pred. No. 1.2e-15;
Matches 76; Conservative 21; Mismatches 59; Indels 10; Gaps 3;

QY 242 KTSGCYGLGMSGVADPQITASSVLE--WTDHTGQENSWKPKKARLKKEG--PPWAAF 297
Db EVNGCSTPLGMEGKIKENKQITASSFKKSGDY-----NEPFRARLNAQGRVNAWQAK 2115
QY 298 ATDEYQWQLDNLKKEKKTGTTGTTGSTMVEHNYVYSAIRILYSDGQKWTVYRPGVEQD 357
Db 2116 ANNKQWLEIDLKIKITAITQCKSLSEMYYKSTIYSEQGVEMKPKYRLKSSMVD 2175

QY 358 KIFGCKNDYHODVNFPLPIIARIRVNPQWQCKIAMKMWELGCG 403
Db 2176 KIFGNTWKGHVKHNFNPPPIISIRVIRVPTKWNQSIIRLELFGC 2221

RESULT 13
MFGM HUMAN
ID MFGM HUMAN STANDARD; PRT; 387 AA.
AC Q08431.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFG-E8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast, and Breast carcinoma;
RX MEDLINE=96213908; PubMed=8639264;
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain.";
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=91371351; PubMed=1909932;
RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
RA Ceriani R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
RN [3]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE=98194924; PubMed=9535276;
RA Guiffrida M.G., Cavaletto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
```

```
RT breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
RN [4]
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=99342076; PubMed=10411933;
RA Haeggqvist B., Naeslund J., Sletten K., Westermarck G.T., Mucchiano G.,
RA Tjernberg L.O., Nordstedt C., Engstrom U., Westermarck P.;
RT "Medin: an integral fragment of aortic smooth muscle cell-produced
RT lactadherin forms the most common human amyloid.";
RN Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97405885; PubMed=9260929;
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD)-dependent cell adhesion.";
RL DNA Cell Biol. 16:861-869(1997).
CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
CC SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
CC -1- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
CC MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC -1- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC AMINO ACID 264 AND 273.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U58516; AAC50549.1; -;
CC EMBL; S58151; AAB19771.1; -;
CC HSPSP; P08709; 1BF9.
CC Genew; HGNC:7036; MFG-E8.
CC MIM; 602281; -;
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR000421; FA58_C.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00181; EGF; 1.
CC SMART; SM00231; FA58C; 2.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT SIGNAL 1 23
FT CHAIN 24 387 LACTADHERIN.
FT CHAIN 202 387 LACTADHERIN; SHORT FORM.
FT CHAIN 268 317 MEDIN.
FT DOMAIN 24 67 EGF-LIKE.
FT DOMAIN 70 225 F5/8 TYPE C 1.
FT DOMAIN 230 387 F5/8 TYPE C 2.
FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 27 38 BY SIMILARITY.
FT DISULFID 32 55 BY SIMILARITY.
FT DISULFID 57 66 BY SIMILARITY.
FT DISULFID 70 225 BY SIMILARITY.
FT DISULFID 212 216 BY SIMILARITY.
FT DISULFID 230 387 BY SIMILARITY.
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 387 AA; 43123 MW; 2EB571DEC83782D CRC64;
```

```
Query Match      8.8%; Score 343; DB 1; Length 387;
Best Local Similarity 41.0%; Pred. No. 2.2e-16;
Matches 71; Conservative 28; Mismatches 68; Indels 6; Gaps 3;

QY 234 LSTSLFTFTKSCYGLTGMESGVADPQITASSVLE-WTDHTGQENSWKPKARLKKPG- 291
DB 218 LRFELGELNCANPLGLKNNISIDPKQITASSYKTWGLHL---FSWNPSYARLDKQGN 274

QY 292 -PPWAAFAATDEYQWLQIDLNKEKITGIITGTSTWVEHNYVYSAVRILYSDGQKWTYVR 350
DB 275 FNAWVAGSYGVNDQWLQDLGSSKEVTGIITQCARNFSGVQFVASYKVAYSNDANWTEYQ 334

QY 351 EPGVEQDKIFQGNKYHODVRNFPPIIARFIRVNPOTQOQKIAMKMLLGC 403
DB 335 DPRTGSSKIFPGNWDNHHKKNLFTPIARYRILPVANHNRILRLLELLGC 387

RESULT 14
MFGM_PIG
ID MFGM_PIG STANDARD; PRT; 409 AA.
AC P79385;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Lactadherin (Milk fat globule-EGF factor 8) (MFG-EB) (MFGM) (Sperm surface protein SP47) (PP47).
GN MFGEB.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RC Tissue-Testis;
SEQUENCE FROM N.A.
RA Ensalin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-BINDING PROTEIN.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND SPERMATOZOAN.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; Y11683; CAA72379.1;
CC HSP; P00740; 1EDM.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR000421; FA58_C.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00231; FA58C; 2.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS01285; FA58C.1; 2.
CC PROSITE; PS01286; FA58C.2; 2.
CC Glycoprotein; Repeat; EGF-like domain.
CC DOMAIN 2 41
CC EGF-LIKE 1.
FT DOMAIN 44 88
FT DOMAIN 91 247
FT DOMAIN 252 409
FT SITE 67 69
FT SITE 6 17
FT DISULFID 6 17
FT DISULFID 11 29
FT DISULFID 31 40

Query Match      8.7%; Score 339; DB 1; Length 409;
Best Local Similarity 41.4%; Pred. No. 4.4e-16;
Matches 72; Conservative 25; Mismatches 69; Indels 8; Gaps 4;

QY 234 LSTSLFTFTKSCYGLTGMESGVADPQITASSVLE-WTDHTG-QENSWKPKARLKKPG 291
DB 240 LRFELGELSCAEPGLKDKNTIPNKQITASSFYRTW---GLSAFWSYPPYARLDNOG 295

QY 292 P--PWAFAATDEYQWLQIDLNKEKITGIITGTSTWVEHNYVYSAVRILYSDGQKWTYV 349
DB 296 KFNWTAQNSASEWLQIDLSQRRTVGITQGARDFGHQIVAAKYAYSDDGVSWTEY 355

QY 350 EPGVEQDKIFQGNKYHODVRNFPPIIARFIRVNPOTQOQKIAMKMLLGC 403
DB 356 RDOGALEKGIFFGNLNDNNHKKMFPETPLTRFVRLPVANHNRITLRLVELLGC 409

RESULT 15
MFGM_RAT
ID MFGM_RAT STANDARD; PRT; 427 AA.
AC P70490;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-EB) (O-acetyl GD3 ganglioside synthase) (AGS) (MFGM).
GN MFGEB OR AGS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC Tissue-Brain;
RX MEDLINE=96374422; PubMed=8780713;
RA Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
RT "Cloning and expression of cDNA for O-acetylation of GD3 ganglioside."
RL Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO PARTICIPATE IN THE O-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; D84068; BAA12210.1;
CC HSP; P00740; 1EDM.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001438; EGF_II.
CC InterPro; IPR000421; FA58_C.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC PRINTS; PR00010; EGFBLD.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00231; FA58C; 2.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS01186; EGF_2; 2.
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:49:45 ; Search time 14.7699 Seconds
(without alignments)
4744.920 Million cell updates/sec

Title: US-10-060-830-3
Perfect score: 3888
Sequence: 1 MFLFLLLLVLLLLLEDAGA.....TQEVSGAGRGDCDVFKEIL 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73; *
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	383	9.9	927	1 JQ0948	A5 antigen precurs
2	361	9.3	2211	1 KFB05	coagulation factor
3	355.5	9.1	2183	2 T42784	coagulation factor
4	351	9.0	2224	1 KFH05	coagulation factor
5	348	9.0	401	2 S65138	glycoprotein antig
6	348	9.0	427	2 S74211	PAS-6/7 protein pr
7	343	8.8	218	2 A47285	milk fat globule p
8	339	8.7	409	2 T11743	pp47 protein - pig
9	339	8.7	427	2 JQ4915	ags protein precur
10	321	8.3	463	1 A36479	milk fat globule m
11	311	8.0	2133	2 T42763	coagulation factor
12	308.5	7.9	216	2 A44258	factor VIII-associ
13	308.5	7.9	2319	2 A47004	-coagulation factor
14	308.5	7.9	2351	1 E2HU	coagulation factor
15	224.5	5.8	845	2 JQ5256	adipocyte transcri
16	216	5.6	1524	2 T30337	polyprotein - Afri
17	214.5	5.5	3623	2 T08618	intrinsic factor-B
18	213	5.5	3133	2 S52093	hemocytin - silkw
19	211	5.4	3623	2 T09456	intrinsic factor-B
20	207.5	5.3	1283	2 T13799	neurexin IV - ful
21	204	5.2	719	2 S51739	transcription repr
22	200.5	5.2	855	2 S42621	protein-tyrosine k
23	198	5.1	277	2 A41735	hyaluronate-bindin
24	198	5.1	707	2 JQ2118	procollagen C-endo
25	195	5.0	823	1 A58788	procollagen C-endo
26	194.5	5.0	730	1 BMH01	procollagen C-endo
27	193	5.0	275	2 JQ6506	tumor necrosis fac
28	193	5.0	1035	1 A43090	enteropeptidase (E
29	192.5	5.0	991	2 I49540	procollagen C-endo

30	192	4.9	1594	2 T30549	hensin - rabbit
31	191.5	4.9	986	1 B58788	procollagen C-endo
32	190.5	4.9	694	2 JQ6554	complement subcomp
33	190	4.9	1464	2 S58984	development protei
34	189.5	4.9	705	1 C1HURB	complement subcomp
35	188.5	4.8	1070	2 T31069	tolloid-BMP-1 like
36	187	4.8	913	2 A48280	receptor tyrosine
37	185.5	4.8	276	2 A47290	TSG-6 homolog p94
38	185.5	4.8	402	2 JH0403	procollagen I C-pr
39	185.5	4.8	688	1 C1HUS	complement subcomp
40	184	4.7	2083	2 T42721	CRP-ductin-alpha p
41	183.5	4.7	449	2 A55362	procollagen I C-pr
42	183	4.7	2403	2 A59386	sanko - human
43	180	4.6	1057	1 A39288	dorsal-ventral pat
44	179	4.6	910	2 A53137	tyrosine kinase re
45	175	4.5	1004	2 T30338	oviductin (EC 3.4.

ALIGNMENTS

RESULT 1

JQ0948
A5 antigen precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: JH0466; JQ0948
R:Takagi, S.; Hirata, T.; Agata, K.; Mochii, M.; Eguchi, G.; Fujisawa, H.
Neuron 7, 295-307, 1991
A:Title: The A5 antigen, a candidate for the neuronal recognition molecule, has homo
A:Reference number: JH0466; MUID:91337458; PMID:1908252
A:Accession: JH0466
A:Molecule type: mRNA
A:Residues: 1-927 <TAK>
A:Cross-references: GB:D10467; GB:D01077; NID:G222962; PIDN:BA01260.1; PID:G222963
A:Experimental source: tadpole, brain
A:Note: This protein has motifs homologous to complement components C1r and C1s and
C:Comment: This protein is a neuronal cell surface molecule involved in the neuronal
C:Superfamily: Xenopus A5 antigen; C1r/C1s repeat homology; discoidin I amino-termi
C:Keywords: duplication; glycoprotein; transmembrane protein
F:1-21/Domain: signal sequence #status predicted <ASig>
F:22-927/Product: A5 antigen #status predicted <ASig>
F:27-138/Domain: C1r/C1s repeat homology <C1r1>
F:147-262/Domain: C1r/C1s repeat homology <C1r2>
F:274-424/Domain: discoidin I amino-terminal homology <DN1>
F:430-584/Domain: discoidin I amino-terminal homology <DN2>
F:646-812/Domain: MAM homology <MAN>
F:861-883/Domain: transmembrane #status predicted <TMM>
F:150,261,300,523,844/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.9%; Score 383; DB 1; Length 927;
Best Local Similarity 25.8%; Pred. No. 3e-19;
Matches 138; Conservative 70; Mismatches 197; Indels 130; Gaps 22;

Qy	5	LLLLLV-----LLLLLEDAGAQGGCGHTVLPSPSGTLTSINYPQTPNSTVCWMEIRV	59
Db	1	MLLRLLSCCCWLLCSLRSSWASRNDKCGDTIKITSPSYLTSAGYPHSPYPPSORCEWLQA	60
Qy	60	KMGERVRIKFG-DFDIEDSDSCHENLYRIYNGICVSRTEICKYKGLGLQNMHS-IESKG	116
Db	61	PEHYQRIIMINPHNPFDELDRE--CKYDYVEIDGDNANGQLLGKYGCG---KIAPSLPSTG	116
Qy	117	NEITLLEFSGIHVSGRGFLASYSVIDKQDLITCLDTASNFLEPFSKYKCPACGLPPFAEI	176
Db	117	PSIFTRFVSDYTPGAGFSIRYEVFK-----TGPE-----CSRNFSS	154
Qy	177	SGTI-----PHGYRDS---SPLCMAGVHAGVY-----SNTLGGQIS-----VVIS	213
Db	155	NGVIKSPKPYEKPNALECTYIIIFAPKMQEIVLEFESFELEADSNAPGQTCRYDWLGIW	214
Qy	214	KGIP-----YYSSSLANNVTSVVGHLSLTFKTSIG-----	245
Db	215	DGFPGVGHIGRYCGQNTPGVRFSFTGLISLMTFHTDSATAKEGFFANFSVQSDTDEDFQ	274

C:Accession: A56172; A42344; A28028; A27498; A25897
 R:Cripe, L.D.; Moore, K.D.; Kane, W.H.
 Biochemistry 31, 3777-3785, 1992

A:Title: Structure of the gene for human coagulation factor V.
 A:Reference number: A42344; MUID:92232668; PMID:1567832

A:Accession: A56172

A:Molecule type: DNA

A:Residues: 1-2224 <CRI>

A:Cross-references: GB:J05368

A:Accession: A42344

A:Molecule type: DNA

A:Residues: 48-5879-89; 120-130; 191-201; 239-249; 313-323; 368-378; 428-437; 461-471; 533-542;
 2070-2111-2120-2172-2181 <CR2>

R:Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kaufm

Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987

A:Title: Complete cDNA and derived amino acid sequence of human factor V.

A:Reference number: A28028; MUID:87260886; PMID:3110773

A:Accession: A28028

A:Molecule type: mRNA

A:Residues: 1-857, 'R', 859-864, 'R', 866-924, 'E', 926-1763, 'I', 1765-2212, 'T', 2214-2224 <JEN>

A:Cross-references: GB:M16967

A:Note: parts of this sequence, including the amino end of the mature protein, were dete

R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.

Biochemistry 26, 6508-6514, 1987

A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum

A:Reference number: A27498; MUID:86107560; PMID:2827731

A:Accession: A27498

A:Molecule type: mRNA

A:Residues: 1-1284, 'I', 1286-1600 <KAN>

A:Cross-references: GB:M17785

A:Note: parts of this sequence were determined by protein sequencing

R:Kane, W.H.; Davie, E.W.

Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986

A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homolog

A:Reference number: A25897; MUID:86313665; PMID:3092220

A:Accession: A25897

A:Molecule type: mRNA

A:Residues: 1188-1215, 1315-2224 <KA2>

A:Cross-references: GB:M14335

A:Note: parts of this sequence were determined by protein sequencing

R:Keller, F.G.; Ortel, T.L.; Quinn-Allen, M.A.; Kane, W.H.

Biochemistry 34, 4118-4124, 1995

A:Title: Thrombin-catalyzed activation of recombinant human factor V.

A:Reference number: A56139; MUID:95210278; PMID:7696276

A:Contents: annotation; thrombin cleavage sites

C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.

C:Genetics:

A:Gene: GDB:F5

A:Cross-references: GDB:119896; OMIM:227400

A:Map position: 1q23-1q23

A:introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65

C:Function:

A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot

A:Pathway: blood coagulation

C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re

C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-2224/Product: coagulation factor V #status predicted <MAT>

F:29-737/Product: coagulation factor Va heavy chain #status experimental <VAH>

F:29-345/Domain: A1 <DA1>

F:33-329/Domain: ferroxidase repeat homology <FO1>

F:346-691/Domain: A2 <DA2>

F:351-684/Domain: ferroxidase repeat homology <FO2>

F:692-1573/Domain: B <DOB>

F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)

F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>

F:1574-1905/Domain: A3 <DA3>

F:51-55,239,297,460,468,554,741,752,760,776,782,821,938,977,1074,1083,1103,1106,1479
 F:167-193,248-329,500-526,603-684,1725-1751,1907-2061,2066-2221/disulfide bonds: #st
 F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
 F:363,693,1546/Binding site: sulfat (Tyr) (covalent) #status predicted
 F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
 F:382,1338/Binding site: carbohydrate (Asn) (covalent) #status absent
 F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted
 F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimen
 F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experim
 F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 9.0%; Score 351; DB 1; Length 2224;

Best Local Similarity 46.4%; Pred. No. 2.3e-16;

Matches 77; Conservative 21; Mismatches 58; Indels 10; Gaps 3;

QY 242 KTSQCYGTIGMESGVADPQITASSVLE--WTDHTGQENSWKPKKARLKKPG--PPWAAF 297

DB 2062 EYNGCSTPLGMENKIKENKQITASSFKKSWGDY-----WEFFARLNAQGRVNAQAK 2115

QY 298 AFDEYQWLOIDLNKKEKKTGTTGTSTWVEHNYYSAYRILYSDDGQKWTYVREPGVEQD 357

DB 2116 ANNNKQWLEIDLLKIKKITAIIQTGCKSLSEMYVKSYTIHYSEQGVENKPYRLKSSMYD 2175

QY 358 KIFQGNKDYHODVRNPLPPIIAREIRYNTQWQKIAMKMLLGC 403

DB 2176 KIFEGTNTKGVHKNFFNPPIISIRFIRVPTKWNOSIALRLEFGC 2221

RESULT 5

S65138

glycoprotein antigen MGP57/53, mammary gland - bovine (fragment)

N:Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-

C:Species: Bos primigenius taurus (cattle)

C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998

C:Accession: S65138; G48394

R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.

Biochim. Biophys. Acta 1245, 385-391, 1995

A:Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal

A:Reference number: S65138; MUID:96125736; PMID:8541316

A:Accession: S65138

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-401 <AK>

R:Mathier, I.H.; Banghart, L.R.; Lane, W.S.

Biochem. Mol. Biol. Int. 29, 545-554, 1993

A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea

II-like sequences.

A:Reference number: A48394; MUID:93250576; PMID:8485470

A:Accession: G48394

A:Status: preliminary

A:Molecule type: protein

A:Residues: 207-220 <MAT>

A:Experimental source: milk

A:Note: sequence extracted from NCBI backbone (NCBIP:131457)

C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF ho

C:Keywords: glycoprotein

F:1-32/Domain: EGF homology (fragment) <EGL>

F:40-79/Domain: EGF homology <EG2>

F:82-239/Domain: discoidin I amino-terminal homology <DN1>

F:243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 9.0%; Score 348; DB 2; Length 401;

Best Local Similarity 42.0%; Pred. No. 2.8e-17;

Matches 73; Conservative 24; Mismatches 69; Indels 8; Gaps 4;

QY 234 LSTSLFTFTSCYGTGLGHESGVADPQITASSVLE-WTDHTG-QENSWKPKKARLKKPG 291

DB 232 LRFELIGCELNGTEPLGLKNDTIPNKQITASSYKTVW---GLSAFWSFPFYARLDNQG 287

QY 292 P--PWAAAFATDEYQWLOIDLNKKEKKTGTTGTSTWVEHNYYSAYRILYSDDGQKWTYV 349

DB 288 KFNWTAQTNSASWLQIDLGSKQKRVTTITQGRDFGHQYVAAYRVAIGDDGVTWTEY 347

JOURNAL Patent: WO 0223660-A 210 21-MAR-2002;
HYSEQ, INC. (US)
FEATURES Location/Qualifiers
source 1..1907
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS 101..1258
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD34902.1"
/db_xref="GI:21439030"
/translation="MFLFLLLLVLLLLDAGAQQDCGHTVLGPESGTLTSINYP
QTPNNTVCWEIRVMKGERVIRKFGDFIEDSDSCHFNLYRIYNGIVSRTEIGKYC
GLQNMNHSIESKNETTLFMSGIHVSGRGLASVIDKQDLTCLDTASFLPEPE
FSKYQAGCLLPAAEISGTPHGYRDSPLCMAGVHGVVNTLGGQISVWISKGIPIY
YESLANNVSYVGHLSLSTLFTKTSKCYGTGLNMGSGRGSSNNITVLEWTDHGO
ENSWKPKKSGNKTALGAFDEYQMDLNLKKEKTIITGTTSTWSTIMCLP
TESCTVMGNGRLCTESLVEQDKIFQGNKRRIITRMVRNPLPPIIARLLE"
BASE COUNT 548 a 449 c 406 g 504 t
ORIGIN

Query Match 100.0%; Score 60; DB 6; Length 1907;
Best Local Similarity 100.0%; Pred. No. 7.4e-26;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGCTGCTCGAGGACGCTGGAGCCAGCAAGTGATGTGGACACACTGTACTAGGC 60
|||||
Db 137 CTGCTGCTCGAGGACGCTGGAGCCAGCAAGTGATGTGGACACACTGTACTAGGC 196
|||||

RESULT 2
AK056350
LOCUS 2020 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens CDNA FLJ31788 fis, clone NT2RI2008598, weakly similar
to MILK FAT GLOBULE-EGF FACTOR 8 PRECURSOR.
ACCESSION AK056350
VERSION 1
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens teratocarcinoma cell line:NT2 CDNA to mRNA,
clone_lib:NT2RI2 clone:NT2RI2008598.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,
Ono, Y., Hotta, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,
Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,
Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,
Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y.,
Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y.,
Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A.,
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2020)
Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp; Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.

FEATURES Location/Qualifiers
source 1..2020
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="NT2RI2008598"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RI2"
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells treated 2-weeks mitotic inhibitor after
5-weeks retinoic acid (RA) induction.-majorily NT2 neuron"
BASE COUNT 494 a 519 c 530 g 477 t
ORIGIN

Query Match 100.0%; Score 60; DB 9; Length 2020;
Best Local Similarity 100.0%; Pred. No. 7.4e-26;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGCTGCTCGAGGACGCTGGAGCCAGCAAGTGATGTGGACACACTGTACTAGGC 60
|||||
Db 531 CTGCTGCTCGAGGACGCTGGAGCCAGCAAGTGATGTGGACACACTGTACTAGGC 590
|||||

RESULT 3
AF387547
LOCUS 2939 bp mRNA linear PRI 11-NOV-2001
DEFINITION Homo sapiens endothelial and smooth muscle cell-derived
neuropilin-like protein (ESDN) mRNA, complete cds.
ACCESSION AF387547
VERSION 1
KEYWORDS AF387547.1 GI:16902434
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2939)
Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,
Matsumori, A., Sasayama, S., Honjo, T. and Tashiro, K.,
ESDN, a novel neuropilin-like membrane protein cloned from vascular
cells with the longest secretory signal sequence among eukaryotes,
is up-regulated after vascular injury
J. Biol. Chem. 276 (36), 34105-34114 (2001),
21424027
PUBMED 11447234

REFERENCE 2 (bases 1 to 2939)
Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,
Sasayama, S., Matsumori, A., Honjo, T. and Tashiro, K.,
Direct Submission
Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto
University Graduate School of Medicine, Yoshida Konoe-cho,
Sakyo-ku, Kyoto 606-8501, Japan
Location/Qualifiers
source 1..2939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
1..2939
/gene="ESDN"
12..2339
/gene="ESDN"
/codon_start=1
/product="endothelial and smooth muscle cell-derived
neuropilin-like protein"
/protein_id="AAL30178.1"
/db_xref="GI:16902435"

FEATURES Location/Qualifiers
source 1..2939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
1..2939
/gene="ESDN"
12..2339
/gene="ESDN"
/codon_start=1
/product="endothelial and smooth muscle cell-derived
neuropilin-like protein"
/protein_id="AAL30178.1"
/db_xref="GI:16902435"

gene
CDS

NDMSGHTTSVGPSTSTFKATGNOPPLVGYNTLLSRTDSCSSAQAQYDTPKAGKP
GLPAPDELIVYQVPOQSVSGRGGEDCVFKEIL"
BASE COUNT 825 a 675 c 657 g 782 t
ORIGIN

Query Match 100.0%; Score 60; DB 9; Length 2939;
Best Local Similarity 100.0%; Pred. No. 7.5e-26;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGCTGCTGAGAGCGCTGAGCCAGCAGGTGATGGATGGACACACTGTACTAGGC 60
Db 186 CTGCTGCTGAGAGCGCTGAGCCAGCAGGTGATGGATGGACACACTGTACTAGGC 245

RESULT 4
LOCUS BC029658 3858 bp mRNA linear PRI 20-MAY-2002
DEFINITION Homo sapiens, similar to endothelial and smooth muscle cell-derived
neuropilin-like protein, clone MGC:30023 IMAGE:4431275, mRNA,
complete cds.
ACCESSION BC029658
VERSION BC029658.1 GI:20988614
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 3858)
Strausberg, R.
Direct Submission
Submitted (06-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: DCTD/DTF
DNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: ang@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 42 Row: c Column: 3.
Location/Qualifiers
1. .3858
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:30023 IMAGE:4431275"
/tissue_type="Prostate, adenocarcinoma."
/clone_lib="NIH_MGC_91"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
21. .2252
/codon_start=1
/products="Similar to endothelial and smooth muscle
cell-derived neuropilin-like protein"
/protein_id="AAH29658.1"
/db_xref="GI:20988615"
/translation="MPLFLILLVLLLLDAGAQGGGHTVLPESGTLRSINYP
QVYNSTCEWEIRVKMGVERVIRKFGDFDIEDSDCHFNLYRIYNGIGVSRTEIGKVC
GLGLONHSTESKNEITLLPWSGIHVSGRGLASVYTDKDLITCLDPSAFLEPPE
FSKYCAPGALLPFAESISGTHPGYRSSPLCMAGVAGVSNVILGQISVIVSKGIPY
YESSLANNVTSVVGHLUSTSLFTFTKSGCYGLMGESGVIAADPOITASSVLEWTDHTGQ
ENSWKPKKARLKPGPWAAFAATDEYQWLIDLNKEKKITGITGTGTMTVEHNYVISA

REMARK
COMMENT
FEATURES
source
1. .3858
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:30023 IMAGE:4431275"
/tissue_type="Prostate, adenocarcinoma."
/clone_lib="NIH_MGC_91"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
21. .2252
/codon_start=1
/products="Similar to endothelial and smooth muscle
cell-derived neuropilin-like protein"
/protein_id="AAH29658.1"
/db_xref="GI:20988615"
/translation="MPLFLILLVLLLLDAGAQGGGHTVLPESGTLRSINYP
QVYNSTCEWEIRVKMGVERVIRKFGDFDIEDSDCHFNLYRIYNGIGVSRTEIGKVC
GLGLONHSTESKNEITLLPWSGIHVSGRGLASVYTDKDLITCLDPSAFLEPPE
FSKYCAPGALLPFAESISGTHPGYRSSPLCMAGVAGVSNVILGQISVIVSKGIPY
YESSLANNVTSVVGHLUSTSLFTFTKSGCYGLMGESGVIAADPOITASSVLEWTDHTGQ
ENSWKPKKARLKPGPWAAFAATDEYQWLIDLNKEKKITGITGTGTMTVEHNYVISA

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 42 Row: c Column: 3.
Location/Qualifiers
1. .3858
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:30023 IMAGE:4431275"
/tissue_type="Prostate, adenocarcinoma."
/clone_lib="NIH_MGC_91"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
21. .2252
/codon_start=1
/products="Similar to endothelial and smooth muscle
cell-derived neuropilin-like protein"
/protein_id="AAH29658.1"
/db_xref="GI:20988615"
/translation="MPLFLILLVLLLLDAGAQGGGHTVLPESGTLRSINYP
QVYNSTCEWEIRVKMGVERVIRKFGDFDIEDSDCHFNLYRIYNGIGVSRTEIGKVC
GLGLONHSTESKNEITLLPWSGIHVSGRGLASVYTDKDLITCLDPSAFLEPPE
FSKYCAPGALLPFAESISGTHPGYRSSPLCMAGVAGVSNVILGQISVIVSKGIPY
YESSLANNVTSVVGHLUSTSLFTFTKSGCYGLMGESGVIAADPOITASSVLEWTDHTGQ
ENSWKPKKARLKPGPWAAFAATDEYQWLIDLNKEKKITGITGTGTMTVEHNYVISA

FEATURES
source

CDS

YRLYSDGQKVTYVREPVEQDKIFQGNKDYHQDVNRNLFPIIARFIRVNPTQMQQ
KIAMELGLGCOIFKRPKLTQPPPNRNSDLAKNTTAPPAKGRAPKFTQPLQPR
SSNEFPAQTEQTASPTDIRNTTVPNTKDVALLVLPVLVMTLTLLILLCACAWH
RNRKKTECTYDLPYWDRAGYLMSLACRHNHGWKMKQFLPAKADVDETPVRYIS
SEVNHLSPREVTYVLOADSAEYQAQPLVGGIVGTLHQRSTFKPESKEAGYADLPYN
SPGQEVHYHAYAEPLFITGPYATPIIMDSMGHPTTSVGQPSSTFKATGNQPPPLVGT
YNTLSRTDSCSSAQAQYDTPKAGKPLPAPDELIVYQVPOQSVSGRGGEDCVFKE
EIL"
BASE COUNT 1137 a 807 c 785 g 1129 t
ORIGIN

Query Match 100.0%; Score 60; DB 9; Length 3858;
Best Local Similarity 100.0%; Pred. No. 7.7e-26;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGCTGCTGAGAGCGCTGAGCCAGCAGGTGATGGATGGACACACTGTACTAGGC 60
Db 57 CTGCTGCTGAGAGCGCTGAGCCAGCAGGTGATGGATGGACACACTGTACTAGGC 116

RESULT 5
LOCUS AB073146 6093 bp mRNA linear PRI 26-APR-2002
DEFINITION Homo sapiens mRNA for CLCPL, complete cds.
ACCESSION AB073146
VERSION AB073146.1 GI:20330503
KEYWORDS
SOURCE Homo sapiens lung cell_line:normal lung tissue cdna to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 6093)
Koshikawa, K., Osada, H., Kozaki, K., Konishi, H., Masuda, A.,
Tatematsu, Y., Mitsudomi, T., Nakao, A. and Takahashi, T.
Significant up-regulation of a novel gene, CLCPL, in a highly
metastatic lung cancer subline as well as in lung cancers in vivo
Oncogene (2002) In press
2 (bases 1 to 6093)
Koshikawa, K., Konishi, H., Osada, H. and Takahashi, T.
Direct Submission
Submitted (16-OCT-2001) Takashi Takahashi, Aichi Cancer Center
Research Institute, Division of Molecular Oncology; Chikusa-ku,
Kanokoden 1-1, Nagoya, Aichi 464-8681, Japan
(E-mail: takaichi-cc.jp, Tel:81-52-764-2993, Fax:81-52-764-2993)

Location/Qualifiers
1. .6093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3q12.1-12.2"
/cell_line="normal lung tissue"
/tissue_type="lung"
1. .6093
/gene="CLCPL"
364. .2691
/gene="CLCPL"
/function="associated with metastatic phenotype"
/note="CUB, ICCL, coagulation factor V/VIII-homology
domains protein 1"
/codon_start=1
/evidence=experimental
/product="CLCPL"
/protein_id="BAB91138.1"
/db_xref="GI:20330504"
/translation="MASRAVVRARRCQCPQVRAAAXAAPAAALPLSRSLPPCSNSS
FSMPLFLILLVLLLEDAGAQGGGHTVLPESGTLRSINYPQYFNSVYCEWE
IRVKMGVERVIRKFGDFDIEDSDCHFNLYRIYNGIGVSRTEIGKICGLGMHNSIES
KNEITLLPWSGIHVSGRGLASVYTDKDLITCLDPSAFLEPPEFSKYCAPGALLP
VGLHSTISLFTKSGCYGLMGESGVIAADPOITASSVLEWTDHTGTGNSKPKKARL
KPGPPWAAFAATDEYQWLIDLNKEKKITGITGTGTMTVEHNYVSAVRLYSDGOKW
VTYREPVEQDKIFQGNKDYHQDVNRNLFPIIARFIRVNPTQMQQKIAMELGLGCO
FIPKGRPPKLTQPPPNRNSDLAKNTTAPPAKGRAPKFTQPLQPRSSNEFPAQTEQT

FEATURES
source

gene
CDS

TASPDIRNTTVPNTKDVLAARVLPVLMVLTLLILVCAWHNRNKKKTEGYD
 LPYWDAGWGMKQKLPKAVDHEETPVYSSSEVNHLSPREVTVLQADSAEYAP
 LVGIVTLHORSTFKPEKGEAGYADLPYNSPGOEYHAYAEPLPTIGPEVATPII
 MDMSGPTTSVGQPSSTFKATGNCQPPPLVGTNTLLSRDSCSSQAQYDTFKAGKP
 GLPAPDELIVQVFSQIEVSGVSGADGECDFKEL

misc_feature

577..924

/gene="CLCP1"

/note="CUB domain"

934..1191

/note="CLCP1"

/gene="CLCP1"

/note="LCCL domain"

1234..1710

/gene="CLCP1"

/note="F5/8 type C domain"

1942..2010

/gene="CLCP1"

/note="transmembrane domain"

4264..4269

/gene="CLCP1"

5873..5878

/gene="CLCP1"

6052..6057

/gene="CLCP1"

BASE COUNT 1671 a 1258 c 1289 g 1874 t 1 others

ORIGIN

Query Match 100.0%; Score 60; DB 9; Length 6093;

Best Local Similarity 100.0%; Pred. No. 7.9e-26; Indels 0; Gaps 0;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGTCGTCGAGGACGCTGGAGCCAGCAGGTGATGATGATGACACACGTACTAGGC 60

|||||

Db 538 CTGTCGTCGAGGACGCTGGAGCCAGCAGGTGATGATGATGACACACGTACTAGGC 597

RESULT 6

AF387548

LOCUS

DEFINITION Mus musculus strain ICR endothelial and smooth muscle cell-derived

neurofilin-like protein (Esdn) mRNA, complete cds.

ACCESSION

AF387548

VERSION

AF387548.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

2642 bp mRNA linear ROD 11-NOV-2001

Mus musculus strain ICR endothelial and smooth muscle cell-derived

neurofilin-like protein (Esdn) mRNA, complete cds.

Mus musculus

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 2642)

Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,

Matsumori, A., Sasayama, S., Honjo, T. and Tashiro, K.

ESDN, a novel neurofilin-like membrane protein cloned from vascular

cells with the longest secretory signal sequence among eukaryotes,

is up-regulated after vascular injury

J. Biol. Chem. 276 (36), 34105-34114 (2001)

21424027

11447234

2 (bases 1 to 2642)

Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N.,

Sasayama, S., Matsumori, A., Honjo, T. and Tashiro, K.

Direct Submission

Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto

University Graduate School of Medicine, Yoshida Konoe-cho,

Sakyo-ku, Kyoto 606-8501, Japan

Location/Qualifiers

1..2642

/organism="Mus musculus"

/strain="ICR"

/db_xref="taxon:10090"

1..2642

/gene="Esdn"

/note="similar to the Mus musculus strain 129/SvJ Esdn

gene sequence in GenBank Accession Number AF442820"

314..2623

CDS

/gene="Esdn"
 /codon_start=1
 /product="endothelial and smooth muscle cell-derived
 neurofilin-like protein"
 /protein_id="AAL30179.1"
 /db_xref="GI:16902437"

misc_feature

431..778

/gene="Esdn"

/note="corresponds to at 453 to 800, referred to in J.

Biol. Chem. 276 (36), 34105-34114 (2001)"

BASE COUNT 639 a 703 c 744 g 556 t

ORIGIN

Query Match 61.78; Score 37; DB 10; Length 2642;

Best Local Similarity 100.0%; Pred. No. 1.1e-11;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 CCACCAAGGTGATGATGATGACACACGTACTAGGC 60

|||||

Db 502 CCACCAAGGTGATGATGATGACACACGTACTAGGC 538

RESULT 7

HSA332516

LOCUS

DEFINITION

Homo sapiens genomic sequence surrounding NotI site, clone

NR5-KF5C.

ACCESSION

AJ332516

VERSION

AJ332516.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

675 bp DNA linear PRI 18-JUL-2002

Homo sapiens genomic sequence surrounding NotI site, clone

NR5-KF5C.

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 675)

Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M.,

Podowski, R.M., Matushkin, Y.G., Gvanchadani, A., Muravenko, O.V.,

Levitsky, V.G., Kolchanov, N.A., Protodanov, A.I., Kashuba, V.I.,

Kisselev, L.L., Wasserman, W., Wahlestedt, C. and Zabarovsky, E.R.

NotI flanking sequences: a tool for gene discovery and verification

of the human genome

Nucleic Acids Res. 30 (14), 3163-3170 (2002)

12136098

2 (bases 1 to 675)

Zabarovsky, E.R.

Direct Submission

Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,

Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,

Sweden

Location/Qualifiers

1..675

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="NR5-KF5C"

125 a 174 c 195 g 180 t 1 others

BASE COUNT

ORIGIN

Query Match 58.3%; Score 35; DB 9; Length 675;

Best Local Similarity 100.0%; Pred. No. 1.7e-10;

[illegible]

```

repeat_region 7226..7437
/rpt_family="MIR"
complement(8514..8674)
/rpt_family="MLT1K"
repeat_region complement(9539..9743)
/rpt_family="MIR"
repeat_region 11368..11494
/rpt_family="MER91C"
repeat_region 11540..11574
/rpt_family="AT_rich"
misc_feature 12184
/feature="HAOL calls T; HCB calls C"
/function="polymorphic site"
complement(14951..14989)
/rpt_family="MER5B"
repeat_region complement(16997..17769)
/rpt_family="L1P11"
repeat_region complement(17770..18078)
/rpt_family="AluY"
repeat_region complement(18079..18787)
/rpt_family="L1P11"
repeat_region complement(18786..20474)
/rpt_family="L1P"
repeat_region complement(20482..20587)
/rpt_family="U6"
repeat_region 20619..20651
/rpt_family="AT_rich"
repeat_region complement(21524..21582)
/rpt_family="MIR"
repeat_region complement(22970..23270)
/rpt_family="AluY"
repeat_region complement(23271..23572)
/rpt_family="AluY"
repeat_region 25217..25528
/rpt_family="AluSp"
repeat_region 25876..26005
/rpt_family="L2"
repeat_region 26112..26173
/rpt_family="TATG)n"
repeat_region 26181..26416
/rpt_family="L2"
repeat_region 27067..27095
/rpt_family="GC_rich"
repeat_region 27773..27806
/rpt_family="GC_rich"
repeat_region 27807..27902
/rpt_family="CT_rich"
repeat_region complement(29172..29330)
/rpt_family="MIR"
repeat_region 29499..29795
/rpt_family="AluSc"
repeat_region 30007..30090
/rpt_family="MIR"
repeat_region 30238..30260
/rpt_family="AT_rich"
repeat_region 30879..30899
/rpt_family="AT_rich"
repeat_region 33832..34102
/rpt_family="L2"

BASE COUNT 10028 a 6968 c 6935 g 11087 t
ORIGIN

Query Match 58.3%; Score 35; DB 9; Length 35018;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTGCTGAGACGCTGGAGCCCGACGAGGTGA 35
|||||
Db 27894 CTGCTGCTGAGACGCTGGAGCCCGACGAGGTGA 27928

RESULT 9
AC025661/c

```

```

LOCUS AC025661 146515 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 3 clone RP11-322A20 map 3, LOW-PASS
SEQUENCE SAMPLING.
AC025661
AC025661.2 GI:7543781
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 146515)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-322A20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 146515)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 146515)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 12, 2000 this sequence version replaced gi:7230271.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smit, A.F.A. & Green, P. (1996-1997)
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

----- Project Information
Center project name: L7942
Center clone name: 322_A_20

* NOTE: This record contains 164 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1
* 836: contig of 836 bp in length
* 837 936: gap of 100 bp
* 937 1748: contig of 812 bp in length
* 1749 1848: gap of 100 bp
* 1849 2680: contig of 832 bp in length
* 2681 2780: gap of 100 bp
* 2781 3534: contig of 754 bp in length
* 3535 3634: gap of 100 bp
* 3635 4383: contig of 749 bp in length
* 4384 4483: gap of 100 bp
* 4484 5261: contig of 778 bp in length
* 5262 5361: gap of 100 bp
* 5362 6136: contig of 775 bp in length
* 6137 6236: gap of 100 bp
* 6237 7045: contig of 809 bp in length
* 7046 7145: gap of 100 bp
* 7146 7953: contig of 808 bp in length
* 7954 8053: gap of 100 bp
* 8054 8857: contig of 804 bp in length
* 8858 8957: gap of 100 bp
* 8958 9784: contig of 827 bp in length
* 9785 9884: gap of 100 bp
* 9885 10718: contig of 834 bp in length
* 10719 10818: gap of 100 bp
* 10819 11640: contig of 822 bp in length
* 11641 11740: gap of 100 bp
* 11741 12534: contig of 794 bp in length
* 12535 12634: gap of 100 bp
* 12635 13407: contig of 773 bp in length
* 13408 13507: gap of 100 bp
* 13508 14276: contig of 769 bp in length
* 14277 14376: gap of 100 bp
* 14377 15185: contig of 809 bp in length
* 15186 15285: gap of 100 bp
* 15286 16086: contig of 801 bp in length
* 16087 16386: gap of 100 bp
* 16387 16986: contig of 800 bp in length
* 16987 17086: gap of 100 bp
* 17087 17913: contig of 827 bp in length
* 17914 18013: gap of 100 bp
* 18014 18841: contig of 828 bp in length
* 18842 18941: gap of 100 bp
* 18942 19770: contig of 829 bp in length
* 19771 19870: gap of 100 bp
* 19871 20689: contig of 819 bp in length
* 20690 20789: gap of 100 bp
* 20790 21575: contig of 786 bp in length
* 21576 21675: gap of 100 bp
* 21676 22423: contig of 748 bp in length
* 22424 22523: gap of 100 bp
* 22524 23324: contig of 799 bp in length
* 23323 23422: gap of 100 bp
* 23423 24239: contig of 817 bp in length
* 24240 24339: gap of 100 bp
* 24340 25199: contig of 860 bp in length
* 25200 25299: gap of 100 bp
* 25300 26131: contig of 832 bp in length
* 26132 26231: gap of 100 bp

* 26232 27103: contig of 872 bp in length
* 27104 27203: gap of 100 bp
* 27204 28060: contig of 857 bp in length
* 28061 28160: gap of 100 bp
* 28161 29013: contig of 853 bp in length
* 29014 29113: gap of 100 bp
* 29114 29944: contig of 831 bp in length
* 29945 30044: gap of 100 bp
* 30045 30746: contig of 702 bp in length
* 30747 30846: gap of 100 bp
* 30847 31652: contig of 806 bp in length
* 31653 31752: gap of 100 bp
* 31753 32559: contig of 807 bp in length
* 32560 32659: gap of 100 bp
* 32660 33470: contig of 811 bp in length
* 33471 33570: gap of 100 bp
* 33571 34386: contig of 816 bp in length
* 34387 34486: gap of 100 bp
* 34487 35290: contig of 804 bp in length
* 35291 35390: gap of 100 bp
* 35391 36199: contig of 809 bp in length
* 36200 36299: gap of 100 bp
* 36300 37156: contig of 857 bp in length
* 37157 37256: gap of 100 bp
* 37257 38097: contig of 841 bp in length
* 38098 38197: gap of 100 bp
* 38198 39015: contig of 818 bp in length
* 39016 39115: gap of 100 bp
* 39116 39895: contig of 780 bp in length
* 39896 39995: gap of 100 bp
* 39996 40786: contig of 791 bp in length
* 40787 40886: gap of 100 bp
* 40887 41686: contig of 800 bp in length
* 41687 41786: gap of 100 bp
* 41787 42549: contig of 763 bp in length
* 42550 42649: gap of 100 bp
* 42650 43459: contig of 810 bp in length
* 43460 43559: gap of 100 bp
* 43560 44293: contig of 734 bp in length
* 44294 44393: gap of 100 bp
* 44394 45225: contig of 832 bp in length
* 45226 45325: gap of 100 bp
* 45326 46138: contig of 813 bp in length
* 46139 46238: gap of 100 bp
* 46239 46987: contig of 749 bp in length
* 46988 47087: gap of 100 bp
* 47088 47853: contig of 766 bp in length
* 47854 47953: gap of 100 bp
* 47954 48758: contig of 805 bp in length
* 48759 48858: gap of 100 bp
* 48859 49664: contig of 806 bp in length
* 49665 49764: gap of 100 bp
* 49765 50569: contig of 805 bp in length
* 50570 50669: gap of 100 bp

Query Match 58.3%; Score 35; DB 2; Length 146515;
Best Local Similarity 100.0%; Pred. No. 2.3e-10;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTGCTCGAGGACGCTGGAGCCAGCAAGGTGA 35
|||||
Db 6442 CTGCTGCTCGAGGACGCTGGAGCCAGCAAGGTGA 6408

RESULT 10
AC091213/c

LOCUS
DEFINITION

AC091213
AC091213

AC091213
AC091213

AC091213
AC091213

AC091213
AC091213

AC091213 Homo sapiens chromosome 3 clone CTD-2011L5, WORKING DRAFT SEQUENCE,
8 unordered pieces.
AC091213
AC091213.13 GI:21539673
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 164216)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Avele,M., Banks,T.,
 Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagb,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
 Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
 Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Thomas,S.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 164216)
 Worley,K.C.
 Direct Submission
 Submitted (04-APR-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 164216)
 Worley,K.C.
 Direct Submission
 Submitted (24-JUN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jun 23, 2002 this sequence version replaced gi:21539104.

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Drafting Center Code: BCM
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: HCIB
 Center clone name: CTD-2011L5

 Summary Statistics

 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 161549 bases at least Q40

Consensus quality: 162327 bases at least Q30
 Consensus quality: 162604 bases at least Q20
 Estimated insert size: 157933; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a working draft sequence. It currently
 consists of 8 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 *
 *
 * 1 3452: contig of 3452 bp in length
 * 3453 3552: gap of unknown length
 * 3553 10549: contig of 6997 bp in length
 * 10550 10649: gap of unknown length
 * 10650 26000: contig of 15351 bp in length
 * 26001 26100: gap of unknown length
 * 26101 44000: contig of 17900 bp in length
 * 44001 58020: contig of 13920 bp in length
 * 58021 58120: gap of unknown length
 * 58121 79698: contig of 21575 bp in length
 * 79699 119936: gap of unknown length
 * 119937 120036: contig of 40141 bp in length
 * 120037 164216: contig of 44180 bp in length.
 *
 * Location/Qualifiers
 1..164216
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clones="CTD-2011L5"
 BASE COUNT 52340 a 31466 c 30520 g 49178 t 712 others
 ORIGIN
 Query Match 58.3%; Score 35; DB 2; Length 164216;
 Best Local Similarity 100.0%; Pred. No. 2.4e-10;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGCTGCTGAGGACGCTGGAGCCGACGAGGTGA 35
 |||||
 Db 131657 CTGCTGCTGAGGACGCTGGAGCCGACGAGGTGA 131623

 RESULT 11
 BC007117
 LOCUS BC007117 586 bp mRNA linear PRI 12-JUL-2001
 DEFINITION Homo sapiens, clone IMAGE:4333276, mRNA, partial cds.
 ACCESSION BC007117
 VERSION BC007117.1 GI:13938000
 KEYWORDS
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 586)
 Strausberg,R.
 Direct Submission
 Submitted (30-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC/DC/DTP
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA sequencing by: Sequencing Group at the Stanford Human Genome

 REMARK
 COMMENT

Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdgapaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 21 Row: n Column: 22.

FEATURES

source
 Location/Qualifiers

1..586
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4333276"
 /tissue_type="Skin, melanotic melanoma, high MDR."
 /clone_lib="NIH_MGC_62"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 <1..106
 /codon_start=2
 /product="Unknown (protein for IMAGE:4333276)"
 /protein_id="AAH07117.1"
 /db_xref="GI:13938001"
 /translation="LLLVLLLLLEPAGAAQVWEARILFAGCAEGHLS"
 190 a 106 c 127 g 163 t

CDS

1..106

BASE COUNT

190 a 106 c 127 g 163 t

ORIGIN

Query Match 53.3%; Score 32; DB 9; Length 586;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

29 AAGGTGATGGATGGACACACTGTACTAGGC 60

|||||
 Db 113 AAGGTGATGGATGGACACACTGTACTAGGC 144

RESULT 12

AF387549

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

12..2319

/gene="Esdn"

12..2321

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

1..2339

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

1..2339

/gene="Esdn"

12..2321

/gene="Esdn"

/codon_start=1

/product="endothelial and smooth muscle cell-derived

neuropilin-like protein"

/protein_id="AAL30180.1"

/db_xref="GI:16902439"

/translation="WASRAPLRAASPDQGGRAAPATGRAPLPSAGWCPLPPGRNS

SSRPRLLLLLLLPDAGAKQDGGHTVLGPESGLTSINYPHYPSNVCCKWEIRV

ETVLFMGISHASGRFLASYSIDKQDLITCLDTVSNFLEPEFSKYCPAGCLLPFAE

ISGTIPHGVRDSSPLCMAGIHAGVYDVLGGIISVYISKGTPTYESSLIANNVSMVGY

LSTSLFTKSCYGLGMSGVADPQITASSVLEWTDHMGQENSWKEKARLRKPG

PPWAFATDEHQLQDLNKKRKITGITVTGLLEHNYYISAYRVLVSDGQKWTIV

REPAAQDKIFQGNKDYHKDVRNFPPIAFIRVNPVQWQOKIAMKVELLGQFTLL

KGRPLKLTQPPPPRNSNLTNTVHPKLGRAKFTQALQPRSRNDLPLPAQTATPDP

VKNTTVPSTVKDVALAIVPVLVMAITLILVLCAMHWRNRKKKAAEGTYDLDPHD

RAGMWKGVKQLLPKAKSVHEETPVYNSVSEVSHLSPREVTTLQADSAEYAOPLVGGI

VGTLHQSTFKPEEKEKASYADLPYNAPVOEVYHAYAEPLPVTGPVATPTVMDMSG

ELVYQVQSTSTFRAGNQPALVGTNTLLSRDSCSSGGAQIDTPKGGKPAAPAE

BASE COUNT 612 a 606 c 594 g 527 t

ORIGIN

Query Match 53.3%; Score 32; DB 10; Length 2339;

Best Local Similarity 100.0%; Pred. No. 1.3e-08;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AAGGTGATGGATGGACACACTGTACTAGGC 60

|||||

Db 205 AAGGTGATGGATGGACACACTGTACTAGGC 236

|||||

AC073443 61344 bp DNA linear HTG 18-JUN-2000

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

AC073443

COMMENT

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L10150
 Center clone name: 147_E_24

 * NOTE: This record contains 75 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 *
 * 1 719: contig of 719 bp in length
 * 820 819: gap of 100 bp
 * 820 1538: contig of 719 bp in length
 * 1539 1638: gap of 100 bp
 * 1639 2338: contig of 700 bp in length
 * 2339 2438: gap of 100 bp
 * 2439 3151: contig of 713 bp in length
 * 3152 3251: gap of 100 bp
 * 3252 3962: contig of 711 bp in length
 * 3963 4062: gap of 100 bp
 * 4063 4780: contig of 718 bp in length
 * 4781 4880: gap of 100 bp
 * 4881 5598: contig of 718 bp in length
 * 5599 5698: gap of 100 bp
 * 5699 6392: contig of 694 bp in length
 * 6393 6492: gap of 100 bp
 * 6493 7209: contig of 717 bp in length
 * 7210 7309: gap of 100 bp
 * 7310 8033: contig of 724 bp in length
 * 8034 8133: gap of 100 bp
 * 8134 8849: contig of 716 bp in length
 * 8850 8949: gap of 100 bp
 * 8950 9676: contig of 727 bp in length
 * 9677 9776: gap of 100 bp
 * 9777 10498: contig of 722 bp in length
 * 10499 10598: gap of 100 bp
 * 10599 11313: contig of 715 bp in length
 * 11314 11413: gap of 100 bp
 * 11414 12120: contig of 707 bp in length
 * 12121 12220: gap of 100 bp
 * 12221 12944: contig of 724 bp in length
 * 12945 13044: gap of 100 bp
 * 13045 13768: contig of 724 bp in length
 * 13769 13868: gap of 100 bp
 * 13869 14557: contig of 889 bp in length
 * 14558 14657: gap of 100 bp
 * 14658 15396: contig of 739 bp in length
 * 15397 15496: gap of 100 bp
 * 15497 16201: contig of 705 bp in length
 * 16202 16301: gap of 100 bp
 * 16302 17027: contig of 726 bp in length
 * 17028 17127: gap of 100 bp
 * 17128 17844: contig of 717 bp in length
 * 17845 17944: gap of 100 bp
 * 17945 19664: contig of 720 bp in length
 * 18665 18764: gap of 100 bp
 * 18765 19489: contig of 725 bp in length
 * 19490 19589: gap of 100 bp
 * 19590 20311: contig of 722 bp in length
 * 20312 20411: gap of 100 bp

20412 21142: contig of 731 bp in length
 * 21143 21242: gap of 100 bp
 * 21243 21947: contig of 705 bp in length
 * 21948 22047: gap of 100 bp
 * 22048 22765: contig of 718 bp in length
 * 22766 22865: gap of 100 bp
 * 22866 23587: contig of 722 bp in length
 * 23588 23687: gap of 100 bp
 * 23688 24391: contig of 704 bp in length
 * 24392 24491: gap of 100 bp
 * 24492 25200: contig of 709 bp in length
 * 25201 25300: gap of 100 bp
 * 25301 26020: contig of 720 bp in length
 * 26021 26120: gap of 100 bp
 * 26121 26856: contig of 736 bp in length
 * 26857 26956: gap of 100 bp
 * 26957 27672: contig of 716 bp in length
 * 27673 27772: gap of 100 bp
 * 27773 28494: contig of 722 bp in length
 * 28495 28594: gap of 100 bp
 * 28595 29322: contig of 728 bp in length
 * 29323 29422: gap of 100 bp
 * 29423 30114: contig of 692 bp in length
 * 30115 30214: gap of 100 bp
 * 30215 30949: contig of 735 bp in length
 * 30950 31049: gap of 100 bp
 * 31050 31787: contig of 738 bp in length
 * 31788 31887: gap of 100 bp
 * 31888 32609: contig of 722 bp in length
 * 32610 32709: gap of 100 bp
 * 32710 33417: contig of 708 bp in length
 * 33418 33517: gap of 100 bp
 * 33518 34260: contig of 743 bp in length
 * 34261 34360: gap of 100 bp
 * 34361 35088: contig of 728 bp in length
 * 35089 35188: gap of 100 bp
 * 35189 35905: contig of 717 bp in length
 * 35906 36005: gap of 100 bp
 * 36006 36725: contig of 720 bp in length
 * 36726 36825: gap of 100 bp
 * 36826 37540: contig of 715 bp in length
 * 37541 37640: gap of 100 bp
 * 37641 38364: contig of 724 bp in length
 * 38365 38464: gap of 100 bp
 * 38465 39297: contig of 733 bp in length
 * 39298 39397: gap of 100 bp
 * 39398 40009: contig of 712 bp in length
 * 40010 40109: gap of 100 bp
 * 40110 40840: contig of 731 bp in length
 * 40841 40940: gap of 100 bp
 * 40941 41665: contig of 725 bp in length
 * 41666 41765: gap of 100 bp
 * 41766 42497: contig of 732 bp in length
 * 42498 42597: gap of 100 bp
 * 42598 43320: contig of 723 bp in length
 * 43321 43420: gap of 100 bp
 * 43421 44140: contig of 720 bp in length
 * 44141 44240: gap of 100 bp
 * 44241 44948: contig of 708 bp in length
 * 44949 45048: gap of 100 bp
 * 45049 45785: contig of 737 bp in length
 * 45786 45885: gap of 100 bp
 * 45886 46587: contig of 702 bp in length
 * 46588 46687: gap of 100 bp
 * 46688 47406: contig of 719 bp in length
 * 47407 47506: gap of 100 bp
 * 47507 48220: contig of 714 bp in length
 * 48221 48320: gap of 100 bp
 * 48321 49049: contig of 729 bp in length
 * 49050 49149: gap of 100 bp
 * 49150 49863: contig of 714 bp in length
 * 49864 49963: gap of 100 bp
 * 49964 50687: contig of 724 bp in length

```
* 50688 50787: gap of 100 bp
* 50788 51513: contig of 726 bp in length
* 51514 51613: gap of 100 bp
* 51614 52348: contig of 735 bp in length
* 52349 52448: gap of 100 bp
* 52449 53173: contig of 725 bp in length
* 53174 53273: gap of 100 bp
* 53274 54001: contig of 728 bp in length
* 54002 54101: gap of 100 bp
* 54102 54828: contig of 727 bp in length
* 54829 54928: gap of 100 bp
* 54929 55628: contig of 700 bp in length
* 55629 55728: gap of 100 bp
* 55729 56450: contig of 722 bp in length
* 56451 56550: gap of 100 bp

Query Match
Best Local Similarity 53.38: Score 32; DB 2; Length 61344;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 29 AAGGTGATGGTGGACACACTGTACTAGGC 60
Db 4264 AAGGTGATGGTGGACACACTGTACTAGGC 4233
|||||
|||||

AC091213 164216 bp DNA linear HTG 24-JUN-2002
AC091213 Homo sapiens chromosome 3 clone CTD-2011L5, WORKING DRAFT SEQUENCE,
8 unordered pieces.
AC091213
AC091213.13 GI:21539673
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 164216)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbala,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,X., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karissom,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,B., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosier,N., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabar,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
```

```

Query Match          53.3%; Score 32; DB 2; Length 164216;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 AAGGTGATGGATGGACACACTGTACTAGGC 60
      |||||
Db 31855 AAGGTGATGGATGGACACACTGTACTAGGC 31886

RESULT 15
AC013497      192247 bp DNA linear HTG 26-MAR-2000
LOCUS Homo sapiens chromosome 3 clone RP11-319J24 map 3, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC013497
AC013497 4 GI:7329299
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 192247)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 3, clone RP11-319J24
Unpublished
2 (bases 1 to 192247)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelia,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardany,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,K., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6692347.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIDR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1852
Center clone name: 319.J.24
----- Summary Statistics
Sequencing vector: M13: M7815; 100% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182870 bases at least Q40
Consensus quality: 187032 bases at least Q30
Consensus quality: 186925 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 191247; sum-of-ctngs
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-ctngs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

```

* as soon as it is available and the accession number will
* be preserved.
* 1 473 572: contig of 472 bp in length
* 573 3945: contig of 3373 bp in length
* 3946 4045: gap of 100 bp
* 4046 11056: contig of 7011 bp in length
* 11057 11156: gap of 100 bp
* 11157 17728: contig of 6572 bp in length
* 17729 17828: gap of 100 bp
* 17829 30556: contig of 12728 bp in length
* 30557 30656: gap of 100 bp
* 30657 49294: contig of 18638 bp in length
* 49295 49394: gap of 100 bp
* 49395 64509: contig of 15115 bp in length
* 64510 64609: gap of 100 bp
* 64610 83042: contig of 18433 bp in length
* 83043 83142: gap of 100 bp
* 83143 112373: contig of 29231 bp in length
* 112374 112473: gap of 100 bp
* 112474 145663: contig of 33190 bp in length
* 145664 145763: gap of 100 bp
* 145764 192247: contig of 46484 bp in length.

```

FEATURES

Source

```

1. 192247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-319J24"
/clone_lib="RPC1-11 Human Male BAC"

```

misc_feature

```

1. 472
/note="assembly_fragment"
clone_end:SP6
vector_side:right"

```

misc_feature

```

573..3945
/note="assembly_fragment"

```

misc_feature

```

4046..11056
/note="assembly_fragment"

```

misc_feature

```

11157..17728
/note="assembly_fragment"

```

misc_feature

```

17829..30556
/note="assembly_fragment"

```

misc_feature

```

30657..49294
/note="assembly_fragment"

```

misc_feature

```

49395..64509
/note="assembly_fragment"

```

misc_feature

```

64610..83042
/note="assembly_fragment"
clone_end:T7
vector_side:left"

```

misc_feature

```

83143..112373
/note="assembly_fragment"

```

misc_feature

```

112474..145663
/note="assembly_fragment"

```

misc_feature

```

145764..192247
/note="assembly_fragment"

```

```

BASE COUNT 58524 a 38157 c 36840 g 57721 t 1005 others
ORIGIN

```

```

Query Match          53.3%; Score 32; DB 2; Length 192247;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 29 AAGGTGATGGATGGACACACTGTACTAGGC 60
      |||||
Db 78872 AAGGTGATGGATGGACACACTGTACTAGGC 78903

```

```

Search completed: January 21, 2003, 13:52:03
Job time : 324.89 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 10:13:26 ; Search time 35.0551 Seconds
(without alignments)
3854.498 Million cell updates/sec

Title: US-10-060-830-1115

Perfect score: 60

Sequence: 1 ctgctgcgcgagcgctgg.....gtggacacactgtactagc 60.

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	100.0	1907	24	Human neurophilin-H
2	60	100.0	1907	24	Human cDNA encoding
3	60	100.0	1907	24	Novel human coding
4	60	100.0	2428	22	Human colon cancer
5	35	58.3	2223	22	CDNA encoding nove
6	32	53.3	568	22	Human foetal liver
7	32	53.3	568	22	Human brain expres
8	32	53.3	568	22	Human bone marrow
9	32	53.3	568	22	Probe #9581 used t

10	32	53.3	568	24	ABS09615	Human genome-deriv
11	29	48.3	228	22	AAV73598	Human foetal liver
12	29	48.3	228	22	AAK22044	Human brain expres
13	29	48.3	228	22	AAK48207	Human bone marrow
14	29	48.3	228	22	AAI54037	Probe #22723 used
15	29	48.3	228	24	ABS22095	Human genome-deriv
16	29	48.3	1388	24	ABK83640	Human cDNA differe
c 17	17	28.3	342	24	ABN61044	Human cancer relat
c 18	17	28.3	1243	23	ABL10943	Drosophila melanog
c 19	17	28.3	1830	23	ABL19783	Drosophila melanog
c 20	17	28.3	2107	22	ABA08602	Human GTP-binding
21	17	28.3	2221	22	ABA19249	Human nervous syst
22	17	28.3	2228	22	ABA19250	Human nervous syst
c 23	17	28.3	3830	23	ABL19782	Drosophila melanog
c 24	17	28.3	4932	23	ABL10942	Drosophila melanog
c 25	17	28.3	6201	23	ABL08032	Drosophila melanog
c 26	16	26.7	273	14	AAQ39786	Expressed Sequence
27	16	26.7	273	14	AAQ59198	Human brain expres
28	16	26.7	326	21	AAQ00375	Rat smooth muscle
29	16	26.7	441	24	ABQ54280	Human ovarian anti
30	16	26.7	1885	21	AAA50339	Human myristoylate
31	16	26.7	1999	11	AAQ04784	Sequence encoding
c 32	16	26.7	2152	22	AAI88753	Human polynucleoti
c 33	16	26.7	2170	22	ABA08666	Human Zn finger pr
c 34	16	26.7	2476	22	AAK94905	Human full-length
35	16	26.7	2589	21	AAA50340	Human myristoylate
36	16	26.7	2589	24	ABK83589	Human cDNA differe
37	16	26.7	2589	24	ABK64392	Human benign prost
38	16	26.7	2666	22	AAH15985	Human cDNA sequenc
c 39	16	26.7	3384	23	AA568268	DNA encoding novel
c 40	16	26.7	4217	21	AAAC7850	Human ORFX ORF2205
41	16	26.7	5342	21	AAAD00374	Rat smooth muscle
42	16	26.7	16687	23	ABL12468	Drosophila melanog
43	16	26.7	16687	23	ABL18110	Drosophila melanog
44	16	26.7	16687	23	ABL18210	Drosophila melanog
c 45	16	26.7	18129	22	AAK64780	Human immune/haema

ALIGNMENTS

RESULT 1

AAAD35994
ID AAD35994 standard; DNA; 1907 BP.

XX
AC AAD35994;

XX
DT 09-AUG-2002 (first entry)

XX
DE Human neurophilin-Hy2 DNA.

XX
KW Human; neurophilin-like polypeptide; neurophilin-Hy2; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopenia; memory; platelet; plastic anaemia; antinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoicide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW cerebrotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW cerebroprotective; neuroprotective; vulnary; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
KW immunosuppressive; chromosome 6q21; gene; ds.

Homo sapiens.

Key Location/Qualifiers
CDS 101..1258
/*tag= a
/product= "Human neurophilin-Hy2 protein"
/note= "This region designated as SEQ.ID.NO:8 is

CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopaenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues and as a food supplement or molecular
CC weight marker. The cDNAs are useful in gene identification, genome
CC mapping, transgenics, as hybridisation probes, for primer design, for
CC gene chips and as a DNA antigen. The present sequence encodes
CC neuropilin-Hy2.
XX
SQ Sequence 1907 BP; 548 A; 449 C; 406 G; 504 T; 0 other;

Query Match 100.0%; Score 60; DB 24; Length 1907;
Best Local Similarity 100.0%; Pred. No. 6e-22;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTCTCGAGGACGCTGGAGCCCAAGGATGGATGGATGGACACACTGTACTAGGC 60
|||||
DB 137 CTGCTCTCGAGGACGCTGGAGCCCAAGGATGGATGGATGGACACACTGTACTAGGC 196

RESULT 3
ABN59799
ID ABN59799 standard; cDNA; 1907 BP.
XX
AC ABN59799;
XX
DT 28-JUN-2002 (first entry)
XX
DE Novel human coding sequence SEQ ID NO: 210.
XX
KW Human; antianaemic; vulnary; antiinflammatory; immunomodulator;
KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US26015.
XX
PR 11-SEP-2000; 2000US-0659671.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI: 2002-292408/33.
DR P-PSDB; ABB97386.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis -
XX
PS Claim 1; SEQ ID NO 210; 509pp; English.
XX
CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibit e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a coding sequence of the
CC invention.
XX
SQ Sequence 1907 BP; 548 A; 449 C; 406 G; 504 T; 0 other;

Query Match 100.0%; Score 60; DB 24; Length 1907;
Best Local Similarity 100.0%; Pred. No. 6e-22;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTCTCGAGGACGCTGGAGCCCAAGGATGGATGGATGGACACACTGTACTAGGC 60
|||||
DB 137 CTGCTCTCGAGGACGCTGGAGCCCAAGGATGGATGGATGGACACACTGTACTAGGC 196

RESULT 4
AAH34855
ID AAH34855 standard; cDNA; 2428 BP.
XX
AC AAH34855;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1937.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
DR P-PSDB; AAG75450.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 3444-3445; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB7789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 2428 BP; 677 A; 560 C; 535 G; 650 T; 6 other;

Query Match 100.0%; Score 60; DB 22; Length 2428;
Best Local Similarity 100.0%; Pred. No. 6e-22;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTCTCGAGGACGCTGGAGCCCAAGGATGGATGGATGGACACACTGTACTAGGC 60

```
|||||
Db 111 CTGCTCGAGGACGCTGGAGCCAGCAAGGTGATGGATGGACACACTGTACTAGGC 170
RESULT 5
AAS41148/c
ID AAS41148 standard; cDNA; 2223 BP.
XX
AC AAS41148;
XX
DT 17-DEC-2001 (first entry)
XX
DE cDNA encoding novel human enzyme polypeptide #364.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ss.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
```

```
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465566/50.
XX P-PSDB; AAU23278.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
XX preventing, treating neural, immune system, muscular, reproductive,
XX pulmonary, cardiovascular, renal, proliferative disorders and cancerous
XX diseases -
XX
XX Claim 4; SEQ ID NO 374; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
XX polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
XX encoding them. The enzyme polypeptides of the invention may comprise the
XX functional classes of oxidoreductases, transferases, hydrolases, lyases,
XX isomerases or ligases. The sequences of the invention are useful in the
XX diagnosis, treatment, prevention and/or prognosis of a wide range of
XX disorders including hyperproliferative disorders (e.g. cancer),
XX immunodeficiency disorders (e.g. AIDS) autoimmune disorders
XX (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
XX metabolic disorders (e.g. phenylketonuria), inflammatory disorders
XX (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
XX blood-related disorders (e.g. haemophilia), reproductive disorders
XX (e.g. infertility) and infectious disorders (e.g. influenza). The
XX polynucleotides of the invention can also be used in gene therapy.
XX AAS40785-AAS41684 represent cDNA sequences encoding for the novel human
XX enzyme polypeptides of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2223 BP; 619 A; 546 C; 441 G; 602 T; 15 other;
XX
XX Query Match 58.3%; Score 35; DB 22; Length 2223;
XX Best Local Similarity 100.0%; Pred. No. 6.8e-09;
XX Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 CTGCTCTCGAGGACGCTGGAGCCGACGAGGTGA 35
XX |||||||||||||||||||||||||||||||
XX Db 2148 CTGCTCTCGAGGACGCTGGAGCCGACGAGGTGA 2114
XX
XX RESULT 6
XX ABA60998
XX ID ABA60998 standard; DNA; 568 BP.
XX
XX AC ABA60998;
XX
XX 01-FEB-2002 (first entry)
XX
XX
```

```
DE Human foetal liver single exon nucleic acid probe #9303.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 1; SEQ ID NO 9303; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 568 BP; 190 A; 83 C; 107 G; 188 T; 0 other;
XX
XX Query Match 53.3%; Score 32; DB 22; Length 568;
XX Best Local Similarity 100.0%; Pred. No. 2.6e-07;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 29 AAGGTGATGGATGTGGACACACTGTACTAGGC 60
XX |||||||||||||||||||||||||||||||
XX Db 231 AAGGTGATGGATGTGGACACACTGTACTAGGC 262
XX
XX RESULT 7
XX AAK09290
XX ID AAK09290 standard; DNA; 568 BP.
XX
XX AC AAK09290;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 9281.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
```

```

XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483446/52.
XX XX
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX PS Example 4; SEQ ID NO: 9281; 650pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is one of the probes of the
XX CC invention.
XX SQ Sequence 568 BP; 190 A; 83 C; 107 G; 188 T; 0 other;

Query Match 53.3%; Score 32; DB 22; Length 568;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 AAGGTGATGGATGGACACACTGTACTAGGC 60
Db 231 AAGGTGATGGATGGACACACTGTACTAGGC 262

RESULT 8
AAK35179
ID AAK35179 standard; DNA; 568 BP.
AC AAK35179;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 9736.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488500/53.
XX XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 25; SEQ ID NO 9581; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 568 BP; 190 A; 83 C; 107 G; 188 T; 0 other;

Query Match 53.3%; Score 32; DB 22; Length 568;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 AAGGTGATGGATGGACACACTGTACTAGGC 60
Db 231 AAGGTGATGGATGGACACACTGTACTAGGC 262

RESULT 9
AAI40895
ID AAI40895 standard; DNA; 568 BP.
XX AC AAI40895;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #9581 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 25; SEQ ID NO 9581; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 568 BP; 190 A; 83 C; 107 G; 188 T; 0 other;

Query Match 53.3%; Score 32; DB 22; Length 568;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 AAGGTGATGGATGGACACACTGTACTAGGC 60
Db 231 AAGGTGATGGATGGACACACTGTACTAGGC 262

```

Best Local Similarity 100.0%; Pred. No. 2.6e-07; Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AAGGTGATGGATGTGGACACACTGTACTAGGC 60
 |||
 Db 231 AAGGTGATGGATGTGGACACACTGTACTAGGC 262

RESULT 10
 ABS09615
 ID ABS09615 standard; DNA; 568 BP.
 AC ABS09615;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human genome-derived single exon.probe from lung SEQ ID No 9606.
 XX
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2002-114183/15.
 DR
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX
 PS Claim 1; SEQ ID No 9606; 634pp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC nucleic acid expressed in the human lung; measuring gene expression to a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method

CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 568 BP; 190 A; 83 C; 107 G; 188 T; 0 other;
 Query Match 53.3%; Score 32; DB 24; Length 568;
 Best Local Similarity 100.0%; Pred. No. 2.6e-07;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 AAGGTGATGGATGTGGACACACTGTACTAGGC 60
 |||
 Db 231 AAGGTGATGGATGTGGACACACTGTACTAGGC 262

RESULT 11
 ABA73598
 ID ABA73598 standard; DNA; 228 BP.
 XX
 AC ABA73598;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 XX Human foetal liver single exon nucleic acid probe #21903.
 DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 KW Homo sapiens.
 OS
 XX
 XX WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-483447/52.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 XX Claim 4; SEQ ID NO 21903; 639pp + sequence listing; English.
 PS
 XX

CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 228 BP; 70 A; 41 C; 54 G; 63 T; 0 other;

Query Match 48.3%; Score 29; DB 22; Length 228;

Best Local Similarity 100.0%; Pred. No. 9.6e-06;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 GTGATGGATGTGGACACACTGTACTAGGC 60
 Db 1 GTGATGGATGTGGACACACTGTACTAGGC 29

RESULT 12

AAK22044
 ID AAK22044 standard; DNA; 228 BP.

XX AC AAK22044;

XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe SEQ ID NO: 22035.

XX KW Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.

XX OS Homo sapiens.

XX PN WO200157275-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000667.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483446/52.

XX PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -

XX PS Example 4; SEQ ID NO: 22035; 650pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.

XX SQ Sequence 228 BP; 70 A; 41 C; 54 G; 63 T; 0 other;

Query Match 48.3%; Score 29; DB 22; Length 228;

Best Local Similarity 100.0%; Pred. No. 9.6e-06;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 GTGATGGATGTGGACACACTGTACTAGGC 60
 Db 1 GTGATGGATGTGGACACACTGTACTAGGC 29

RESULT 13

AAK48207
 ID AAK48207 standard; DNA; 228 BP.

XX AC AAK48207;

XX DT 06-NOV-2001 (first entry)

XX DE Human bone marrow expressed single exon probe SEQ ID NO: 22764.

XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX OS Homo sapiens.

XX PN WO200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000668.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488900/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -

XX PS Example 4; SEQ ID NO: 22764; 658pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.

XX SQ Sequence 228 BP; 70 A; 41 C; 54 G; 63 T; 0 other;

Query Match 48.3%; Score 29; DB 22; Length 228;

Best Local Similarity 100.0%; Pred. No. 9.6e-06;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 GTGATGGATGTGGACACACTGTACTAGGC 60
 Db 1 GTGATGGATGTGGACACACTGTACTAGGC 29

RESULT 14

AAI54037
 ID AAI54037 standard; DNA; 228 BP.

XX AC AAI54037;

XX DT 17-OCT-2001 (first entry)

```
XX DE Probe #22723 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 25; SEQ ID No 22723; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 228 BP; 70 A; 41 C; 54 G; 63 T; 0 other;

Query Match 48.3%; Score 29; DB 22; Length 228;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 GTGATGGATGTGGACACACTGTACTAGGC 60
Db 1 GTGATGGATGTGGACACACTGTACTAGGC 29

RESULT 15
ABS22095
ID ABS22095 standard; DNA; 228 BP.
XX AC ABS22095;
XX 19-AUG-2002 (first entry)
XX Human genome-derived single exon probe ORF from lung SEQ ID No 22086.
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
XX Chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary hypertension;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease; open reading frame; ORF.
XX OS Homo sapiens.
XX PN WO200186003-A2.
```

```
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US00665.
XX PR 04-FEB-2000; 2000US-180312P.
XX PR 26-MAY-2000; 2000US-207456P.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-234687P.
XX PR 27-SEP-2000; 2000US-236359P.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX DR Spatially-addressable set of single exon nucleic acid probes, used to
XX PT measure gene expression in human lung samples -
XX PS Claim 4; SEQ ID No 22086; 634pp; English.
XX CC The invention relates to a spatially-addressable set of single exon
XX CC nucleic acid probes for measuring gene expression in a sample derived
XX CC from human lung comprising single exon nucleic acid probes having one of
XX CC 12614 nucleic acid sequences mentioned in the specification, or their
XX CC complements or the 12614 open reading frames derived from the 12614
XX CC probes. Also included are a microarray comprising the novel set of
XX CC probes; the novel set of probes which hybridise at high stringency to a
XX CC nucleic acid expressed in the human lung; measuring gene expression in a
XX CC sample derived from human lung, comprising (a) contacting the array with
XX CC a collection of detectably labeled nucleic acids derived from human lung
XX CC mRNA, and (b) measuring the label detectably bound to each probe of
XX CC the array; identifying exons in a eukaryotic genome, comprising
XX CC (a) algorithmically predicting at least one exon from genomic sequences
XX CC of the eukaryote; and (b) detecting specific hybridisation of detectably
XX CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX CC having a fragment identical to the predicted exon, the probe is included
XX CC in the above mentioned microarray; assigning exons to a single gene,
XX CC comprising (a) identifying exons from genomic sequence by the method
XX CC above and (b) measuring the expression of each of the exons in several
XX CC tissues and/or cell types using hybridisation to a single exon
XX CC microarrays having a probe with the exon, where a common pattern of
XX CC expression of the exons in the tissues and/or cell types indicates that
XX CC the exons should be assigned to a single gene; a peptide comprising one
XX CC of 12011 sequences, mentioned in the specification, or encoded by the
XX CC probes/open reading frames (ORF). The probes are used for gene
XX CC expression analysis, and for identifying exons in a gene, particularly
XX CC using human lung derived mRNA and for the study of lung diseases
XX CC such as asthma, lung cancer, chronic obstructive pulmonary disease
XX CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
XX CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
XX CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
XX CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
XX CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
XX CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
XX CC and hyaline membrane disease. The present sequence is a single exon
XX CC probe open reading frame of the invention.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 228 BP; 70 A; 41 C; 54 G; 63 T; 0 other;

Query Match 48.3%; Score 29; DB 24; Length 228;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 GTGATGGATGTGGACACACTGTACTAGGC 60
|||||
|||||
```

Db 1 GTGATGGATGTGGACACACTGTACTAGGC 29

Search completed: January 21, 2003, 13:49:59
Job time : 37.0551 secs

CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/406.030A
FILING DATE: 17-MAR-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/243,391
FILING DATE: 13-MAY-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/985,586
FILING DATE: 03-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/911,533
FILING DATE: 10-JUL-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/787,840
FILING DATE: 05-NOV-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/789,188
FILING DATE: 05-NOV-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11704
FILING DATE: 02-DEC-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/09627
FILING DATE: 05-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: TKT95-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 810 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 471..810
US-08-406-030A-18

Query Match 25.0%; Score 15; DB 4; Length 810;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GAGGACGCTGGAGCC 24
DB 250 GAGGACGCTGGAGCC 236
|||||

RESULT 3
US-09-150-900-48
Sequence 48, Application US/09150900
Patent No. 6207425
GENERAL INFORMATION:
APPLICANT: Liu, Qulang
APPLICANT: Sommer, Steve S.
TITLE OF INVENTION: BIDIRECTIONAL PCR AMPLIFICATION OF SPECIFIC ALLELES
FILE REFERENCE: BI-PASA
CURRENT APPLICATION NUMBER: US/09/150.900
CURRENT FILING DATE: 1998-09-10

EARLIER APPLICATION NUMBER: 60/058575
EARLIER FILING DATE: 1997-09-11
NUMBER OF SEQ ID NOS: 48
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 48
LENGTH: 820
TYPE: DNA
ORGANISM: Homo sapiens
US-09-150-900-48

Query Match 25.0%; Score 15; DB 4; Length 820;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GATGTGGACACACTG 52
DB 725 GATGTGGACACACTG 739
|||||

RESULT 4
US-08-634-331-1
Sequence 1, Application US/08634331
Patent No. 5707809
GENERAL INFORMATION:
APPLICANT: HALVERSON, Joy L.
APPLICANT: DVORAK, Jan
TITLE OF INVENTION: AVIAN SEX IDENTIFICATION PROBES
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634.331
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SHERWOOD, Pamela J.
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: A-55362-3/BIR/PJS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 2777299FHT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 959 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-634-331-1

Query Match 25.0%; Score 15; DB 1; Length 959;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 GCTGGAGCCAGCAA 30
DB 664 GCTGGAGCCAGCAA 678
|||||

RESULT 5
5508165-1
Patent No. 5508165

APPLICANT: HALVERSON, JOY L.; DVORAK, JAN
TITLE OF INVENTION: AVIAN SEX DETERMINATION PROBE
NUMBER OF SEQUENCES: 1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,131
FILING DATE: 09-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 947,100
FILING DATE: 17-SEP-1992
APPLICATION NUMBER: 585,915
FILING DATE: 21-SEP-1990
SEQ ID NO: 1:
LENGTH: 959
5508165-1

Query Match 25.0%; Score 15; DB 6; Length 959;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 GCTGGAGCCCGAGCAA 30
Db 664 GCTGGAGCCCGAGCAA 678

RESULT 6

US-08-356-405-8/c
Sequence 8, Application US/08356405
Patent No. 5807691
GENERAL INFORMATION:
APPLICANT: Amlaiky, No. 5807691rdline
APPLICANT: Boschert, Ursula
APPLICANT: Hen, Rene.
APPLICANT: Plasset, Jean-Luc
TITLE OF INVENTION: Polypeptides Having Serotonin Receptor
TITLE OF INVENTION: Activity (5HT_{2A}), Nucleic Acids Coding for These
TITLE OF INVENTION: Polypeptides and Uses Thereof
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhone-Poulenc Rorer Inc.
STREET: 500 Arcola Road, 3C43
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19426-0107
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,405
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR93/00650
FILING DATE: 29-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/08081
FILING DATE: 01-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Julie K
REGISTRATION NUMBER: 38,619
REFERENCE/DOCKET NUMBER: EX92004-US
TELEPHONE: (610)454-3839
TELEFAX: (610)454-3808
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1071
US-08-356-405-8

Query Match 25.0%; Score 15; DB 1; Length 1073;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 AGCCGACCAAGGTGA 35
Db 145 AGCCGACCAAGGTGA 131

RESULT 7

US-09-067-626-1
Sequence 1, Application US/09067626
Patent No. 6177086
GENERAL INFORMATION:
APPLICANT: Riley, Lee W.
APPLICANT: Nathan, Carl F.
APPLICANT: Ehit, Sabine
TITLE OF INVENTION: DNA MOLECULE CONFERRING ON MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS RESISTANCE AGAINST ANTIMICROBIAL REACTIVE
TITLE OF INVENTION: OXYGEN AND NITROGEN INTERMEDIATES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/067,626
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/045,688
FILING DATE: 06-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/491
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1244 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-067-626-1

Query Match 25.0%; Score 15; DB 4; Length 1244;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GCTCGAGGACGCTGG 20
Db 326 GCTCGAGGACGCTGG 340

RESULT 8

US-09-067-626-3/c
; Sequence 3, Application US/09067626
; Patent No. 6177086
; GENERAL INFORMATION:
; APPLICANT: Riley, Lee W.
; APPLICANT: Nathan, Carl F.
; APPLICANT: Ehrt, Sabine
; TITLE OF INVENTION: DNA MOLECULE CONFERRING ON MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS RESISTANCE AGAINST ANTIMICROBIAL REACTIVE
; TITLE OF INVENTION: OXYGEN AND NITROGEN INTERMEDIATES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/067,626
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/045,688
; FILING DATE: 06-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-067-626-3
Query Match 25.0%; Score 15; DB 4; Length 1244;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 GCTCGAGGACGCTGG 20
DB 919 GCTCGAGGACGCTGG 905
RESULT 9
US-08-634-331-2
; Sequence 2, Application US/08634331
; Patent No. 5707809
; GENERAL INFORMATION:
; APPLICANT: HALVERSON, Joy L.
; APPLICANT: DVORAK, Jan
; TITLE OF INVENTION: AVIAN SEX IDENTIFICATION PROBES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/634,331
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: A-55362-3/BIR/PJS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277729PHT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-634-331-2
Query Match 25.0%; Score 15; DB 1; Length 1899;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 GCTGGAGCCCGAGCA 30
DB 1604 GCTGGAGCCCGAGCA 1618
RESULT 10
US-09-227-717-3
; Sequence 3, Application US/09227717
; Patent No. 6239268
; GENERAL INFORMATION:
; APPLICANT: Lovenberg, Timothy W.
; APPLICANT: Oltersdorf, Tilman
; APPLICANT: Liaw, Chen W.
; APPLICANT: Clevenger, William R.
; APPLICANT: Desouza, Eriol B.
; TITLE OF INVENTION: INTERLEUKIN-1 TYPE 3 RECEPTORS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,717
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/526,704
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 690068.402C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836
; INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 2044 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 89..1771
US-09-227-717-3

Query Match 25.0%; Score 15; DB 4; Length 2044;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GATGTGGACACACTG 52

|||||

DB 1295 GATGTGGACACACTG 1309

RESULT 11

US-08-790-374-1
Sequence 1, Application US/08790374
Patent No. 5863734

GENERAL INFORMATION:
APPLICANT: Karaylorgou, Maria
APPLICANT: Gogos, Joseph A.
TITLE OF INVENTION: METHODS OF TREATMENT FOR OBSESSIVE-COMPULSIVE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/790,374
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-189
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3651 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
US-08-790-374-1

Query Match 25.0%; Score 15; DB 2; Length 3651;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GATGTGGACACACTG 52

|||||

DB 2303 GATGTGGACACACTG 2317

RESULT 12

US-09-103-840A-2

Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, John C.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 25.0%; Score 15; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CTCGAGGACGCTGGA 21

|||||

DB 2596042 CTCGAGGACGCTGGA 2596056

RESULT 13

US-08-592-126-115
Sequence 115, Application US/08592126
Patent No. 5821091

GENERAL INFORMATION:
APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
TITLE OF INVENTION: Polypeptides
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,126
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 115:
SEQUENCE CHARACTERISTICS:

LENGTH: 283 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: H989.seq
US-08-592-126-115

Query Match 23.3%; Score 14; DB 1; Length 283;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 TGGAGCCCGCAG 31
| | | | | | | | | |
Db 197 TGGAGCCCGCAG 210

RESULT 14

US-09-060-756-67
; Sequence 67, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; CURRENT FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 67
; LENGTH: 434
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-060-756-67

Query Match 23.3%; Score 14; DB 4; Length 434;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 GCTCGAGGACGCTG 19
| | | | | | | | | |
Db 186 GCTCGAGGACGCTG 199

RESULT 15

US-09-740-235-9
; Sequence 9, Application US/09740235
; Patent No. 6448041
; GENERAL INFORMATION:
; APPLICANT: Wolven, Amy K.
; APPLICANT: Krasnow, Randi E.
; APPLICANT: Warren, Bridget A.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: COLON CANCER MARKER
; FILE REFERENCE: PC-0031 CIP
; CURRENT APPLICATION NUMBER: US/09/740,235
; CURRENT FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6448041 6923049H1
US-09-740-235-9

Query Match 23.3%; Score 14; DB 4; Length 556;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GCTGGAGCCCGCAG 29
| | | | | | | | | |
Db 394 GCTGGAGCCCGCAG 407

Search completed: January 21, 2003, 15:09:33
Job time : 535.654 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 13:00:06 ; Search time 8.12598 Seconds
(without alignments)
3292.167 Million cell updates/sec

Title: US-10-060-830-1115

Perfect score: 60

Sequence: 1 CTCTGCTCGAGGACGCTGG.....gtggacacactgtactagc 60

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 393868 seqs, 222934149 residues

Word size : 0

Total number of hits satisfying chosen parameters: 787736

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	60	100.0	5657	US-09-974-298-96	Sequence 96, Appl
2	32	53.3	568	US-09-864-761-12562	Sequence 12562, A
3	29	48.3	228	US-09-864-761-29163	Sequence 29163, A
4	16	26.7	397	US-09-960-352-11475	Sequence 11475, A
5	15	25.0	272	US-09-294-093B-4113	Sequence 4113, Ap
6	15	25.0	293	US-09-923-876-3983	Sequence 3983, Ap
7	15	25.0	401	US-09-946-807-1148	Sequence 1148, Ap
8	15	25.0	401	US-09-946-807-1149	Sequence 1149, Ap
9	15	25.0	401	US-09-946-807-1530	Sequence 1530, Ap
10	15	25.0	401	US-09-795-668-1148	Sequence 1148, Ap
11	15	25.0	401	US-09-795-668-1149	Sequence 1149, Ap
12	15	25.0	401	US-09-795-668-1530	Sequence 1530, Ap
13	15	25.0	401	US-09-795-668-1148	Sequence 1148, Ap
14	15	25.0	401	US-09-795-668-1149	Sequence 1149, Ap
15	15	25.0	401	US-09-795-668-1530	Sequence 1530, Ap
16	15	25.0	444	US-09-736-457-465	Sequence 465, App
17	15	25.0	444	US-09-902-941-465	Sequence 465, App
18	15	25.0	444	US-09-849-626-465	Sequence 465, App
19	15	25.0	569	US-09-864-761-11903	Sequence 11903, A

20	15	25.0	628	10	US-09-816-523-16	Sequence 16, Appl
21	15	25.0	960	10	US-09-886-055-388	Sequence 388, App
c 22	15	25.0	1327	12	US-10-044-090-642	Sequence 642, App
23	15	25.0	1350	10	US-09-925-301-266	Sequence 266, App
24	15	25.0	1365	9	US-09-894-844-73	Sequence 73, Appl
25	15	25.0	1686	10	US-09-775-046-14	Sequence 14, Appl
26	15	25.0	1868	9	US-09-981-353-18	Sequence 18, Appl
27	15	25.0	2156	9	US-09-954-531-1362	Sequence 1362, Ap
c 28	15	25.0	2248	10	US-09-925-299-107	Sequence 107, App
c 29	15	25.0	2436	10	US-09-815-242-7778	Sequence 7778, Ap
30	15	25.0	2718	10	US-09-905-983-57	Sequence 57, Appl
31	15	25.0	3651	10	US-09-880-107-3870	Sequence 3870, Ap
32	15	25.0	45237	10	US-09-933-267A-1	Sequence 1, Appl
33	15	25.0	1503841	9	US-09-946-807-1	Sequence 1, Appl
34	15	25.0	1503841	10	US-09-795-668-1	Sequence 1, Appl
c 35	15	25.0	1503841	10	US-09-795-668-1	Sequence 25055, A
36	14	23.3	91	10	US-09-864-761-23111	Sequence 23111, A
37	14	23.3	178	10	US-09-864-761-27790	Sequence 27790, A
c 38	14	23.3	186	10	US-09-864-761-27790	Sequence 5345, Ap
39	14	23.3	211	10	US-09-294-093B-5345	Sequence 4953, Ap
40	14	23.3	224	10	US-09-923-876-4953	Sequence 4953, Ap
41	14	23.3	241	10	US-09-983-965-5672	Sequence 5672, Ap
42	14	23.3	257	10	US-09-923-876-5672	Sequence 9039, Ap
43	14	23.3	371	9	US-09-796-692-9039	Sequence 9039, Ap
c 44	14	23.3	455	10	US-09-864-761-2223	Sequence 2223, Ap
c 45	14	23.3	461	10	US-09-864-761-11806	Sequence 11806, A

ALIGNMENTS

RESULT 1

US-09-974-298-96
; Sequence 96, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hwei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 96
; LENGTH: 5657
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 241227.17
; NAME/KEY: unsure
; LOCATION: 4516
; OTHER INFORMATION: a, t, c, g, or other
US-09-974-298-96

Query Match 100.0%; Score 60; DB 9; Length 5657;

Best Local Similarity 100.0%; Pred No. 2.4e-23;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCTGCTCGAGGACGCTGGAGCCCGACAGGTGATGGATGTGGACACACTGTACTAGGC 60

|||||

Db 37 CTCTGCTCGAGGACGCTGGAGCCCGACAGGTGATGGATGTGGACACACTGTACTAGGC 96

RESULT 2

US-09-864-761-12562

; Sequence 12562, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 29163
LENGTH: 228
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC013497.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
OTHER INFORMATION: NT HIT: D29810.1, EVALUE 1.00e-125
OTHER INFORMATION: EST HUMAN HIT: A1565996.1, EVALUE 1.00e-125
OTHER INFORMATION: SWISSPROT HIT: P98063, EVALUE 1.00e-09

Query Match 48.3%; Score 29; DB 10; Length 228;
Best Local Similarity 100.0%; Pred.No. 9.5e-07;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 GTGATGGATGGGCACACTGTACTAGGC 60
|||||
DB 1 GTGATGGATGGGCACACTGTACTAGGC 29
|||||

```
RESULT 4
US-09-960-352-11475
; Sequence 11475, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11475
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 49-LIB34-003-Q1-E1-E2
US-09-960-352-11475
Query Match 26.7%; Score 16; DB 10; Length 397;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 GCTGGAGCCGACGACG 31
DB 318 GCTGGAGCCGACGACG 333

RESULT 5
US-09-294-093B-4113/C
; Sequence 4113, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 4113
; LENGTH: 272
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700353144H1
US-09-294-093B-4113
Query Match 25.0%; Score 15; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40
DB 173 AGCAAGGTGATGGAT 159

RESULT 6
US-09-923-876-3983/C
; Sequence 3983, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ico)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 3983
; LENGTH: 293
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700454202H1
US-09-923-876-3983
Query Match 25.0%; Score 15; DB 10; Length 293;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40
DB 229 AGCAAGGTGATGGAT 215

RESULT 7
US-09-946-807-1148
; Sequence 1148, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345 2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1148
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-09-946-807-1148
Query Match 25.0%; Score 15; DB 9; Length 401;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40
DB 249 AGCAAGGTGATGGAT 263

RESULT 8
US-09-946-807-1149
; Sequence 1149, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
```

; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946.807
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795.668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US/09/515.716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1149
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-946-807-1149

Query Match 25.0%; Score 15; DB 9; Length 401;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40
|||||
Db 11 AGCAAGGTGATGGAT 25

RESULT 9
US-09-946-807-1530
; Sequence 1530, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946.807
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795.668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US/09/515.716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1530
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-946-807-1530

Query Match 25.0%; Score 15; DB 9; Length 401;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40
|||||
Db 298 AGCAAGGTGATGGAT 312

RESULT 10
US-09-795-668-1148
; Sequence 1148, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795.668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US/09/515.716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1148
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-668-1148

Query Match 25.0%; Score 15; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40
|||||
Db 249 AGCAAGGTGATGGAT 263

RESULT 11
US-09-795-668-1149
; Sequence 1149, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795.668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US/09/515.716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1149
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-668-1149

Query Match 25.0%; Score 15; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40
|||||
Db 11 AGCAAGGTGATGGAT 25

RESULT 12
US-09-795-668-1530
; Sequence 1530, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795.668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US/09/515.716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1530
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-668-1530

Query Match 25.0%; Score 15; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40

Db 298 AGCAAGGTGATGGAT 312
|||||

RESULT 13

US-09-795-686-1148
; Sequence 1148, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1148
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-686-1148

Query Match 25.0%; Score 15; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40
|||||
Db 249 AGCAAGGTGATGGAT 263

RESULT 14

US-09-795-686-1149
; Sequence 1149, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1149
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-686-1149

Query Match 25.0%; Score 15; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40
|||||
Db 11 AGCAAGGTGATGGAT 25

RESULT 15

US-09-795-686-1530
; Sequence 1530, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur

; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1530
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-686-1530

Query Match 25.0%; Score 15; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AGCAAGGTGATGGAT 40
|||||
Db 298 AGCAAGGTGATGGAT 312

Search completed: January 21, 2003, 14:46:48
Job time : 11.126 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 21, 2003, 13:50:07 ; Search time 6.4252 Seconds
(without alignments)
2488.651 Million cell updates/sec
Title: US-10-060-830-1115
Perfect score: 20
Sequence: 1 ctgctgctgagcagctgg.....gtggacacactgtactagc 60

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 908470 seqs, 133250620 residues
Word size: 1
Total number of hits satisfying chosen parameters: 1687399

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10060830/runat_16012003_092701_1480/app_query.fasta_1.1109
-DB=A_Geneseq_101002 -QFT=fastan -SUFFIX=olin2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn.1.1.36 @runat_16012003_092701_1480 -NCPU=3
-NO_XLPXY -NO_WMAP -LARGOQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002: *
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: *
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: *
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: *
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: *
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: *
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: *
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: *
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: *
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: *
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: *
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: *
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: *
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: *
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: *
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: *
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: *
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: *
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: *
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: *
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: *
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: *
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: *
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	385	23	Human neuropilin-H
2	20	100.0	385	23	Human neuropilin-H
3	20	100.0	385	23	Novel human protei
4	20	100.0	385	23	Human colon cancer
5	12	60.0	365	22	Human neuropilin-H
6	9	45.0	75	22	Peptide #9723 enco
7	9	45.0	75	22	Human brain expres
8	9	45.0	75	22	Human bone marrow
9	9	45.0	75	22	Peptide #10061 enc
10	9	45.0	75	22	Human peptide enco
11	8	40.0	75	23	Human neuropilin-H
12	7	36.8	22	21	Human cancer assoc
13	7	36.8	36	22	Novel human diagno
14	7	36.8	59	21	Human secreted pro
15	7	36.8	76	22	Human foetal prote
16	7	36.8	84	21	Arabidopsis thalia
17	7	36.8	85	22	Human polypeptide
18	7	35.0	86	23	Human OREF protein
19	7	36.8	87	21	Arabidopsis thalia
20	7	36.8	89	21	Arabidopsis thalia
21	7	36.8	91	22	Human polypeptide
22	7	36.8	106	21	Arabidopsis thalia
23	7	36.8	106	22	Novel human diagno
24	7	36.8	128	21	Human secreted pro
25	7	36.8	141	21	Arabidopsis thalia
26	7	36.8	143	21	Arabidopsis thalia
27	7	36.8	143	22	A. thaliana psaki
28	7	36.8	160	21	Arabidopsis thalia
29	7	36.8	187	20	Human normal bladd
30	7	36.8	268	21	Arabidopsis thalia
31	7	36.8	268	21	Arabidopsis thalia
32	7	36.8	297	22	Drosophila melanog
33	7	36.8	310	21	Arabidopsis thalia
34	7	36.8	310	21	Arabidopsis thalia
35	7	36.8	343	22	Human protein SEQ
36	7	36.8	361	22	Novel human diagno
37	7	36.8	422	20	Enterococcus faeca
38	7	36.8	422	23	E faecalis F129 a
39	7	36.8	424	22	Drosophila melanog
40	7	36.8	449	20	Enterococcus faeca
41	7	36.8	449	23	E faecalis F129 p
42	7	36.8	450	21	Arabidopsis thalia
43	7	36.8	463	23	Novel human protei
44	7	36.8	479	22	Drosophila melanog
45	7	36.8	484	21	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAE22716
ID AAE22716 standard; Protein; 385 AA.
XX
XX
AC AAE22716;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy2 protein.

XX Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoicide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;

CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX SQ Sequence 583 AA;

Alignment Scores:
 Pred. No.: 9.54e-13 Length: 583
 Score: 20.00 Matches: 20
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-10-060-830-1115 (1-60) x AAG75450 (1-583)

Qy 1 CTGCTGCTCGAGGAGCTGGAGCCAGCAAGTGATGTGGACACACTGTACTAGGC 60

Db 37 LeuLeuLeuGluAspAlaGlyAlaGlnGlnGlyAspGlyCysGlyHisThrValLeuGly 56

RESULT 5

AAE22721

ID AAE22721 standard; Protein: 365 AA.

XX AC AAE22721;

XX DT 09-AUG-2002 (first entry)

XX DE Human neuropilin-Hy2 mature protein.

XX KW Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopenia; memory; platelet; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW nontropic; neuroprotective; vulnarary; anticonvulsant; antiparastic;
 KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
 KW immunosuppressive.

XX OS Homo sapiens.

XX PN WO200222815-A1.

XX PD 21-MAR-2002.

XX PF 12-SEP-2001; 2001WO-US28488.

XX PR 11-SEP-2000; 2000US-0659671.

XX PR 06-SEP-2001; 2001US-317902P.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT;

XX DR WPI; 2002-393966/42.

XX PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects
 XX PS Disclosure; Page 131-132; 152pp; English.

XX CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and

CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration. In wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy2
 CC mature protein.

XX SQ Sequence 365 AA;

Alignment Scores:

Pred. No.: 0.000224 Length: 365
 Score: 12.00 Matches: 12
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 60.00% Indels: 0
 DB: 23 Gaps: 0

US-10-060-830-1115 (1-60) x AAE22721 (1-365)

Qy 25 CAGCAAGTGATGTGGACACACTGTACTAGGC 60

Db 1 GlnGlnGlyAspGlyCysGlyHisThrValLeuGly 12

RESULT 6

ABBA42217

ID ABB42217 standard; Peptide: 75 AA.

XX AC ABB42217;

XX DT 04-FEB-2002 (first entry)

XX DE Peptide #9723 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX PN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00669.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 XX Claim 27; SEQ ID NO 34852; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 75 AA;
 SQ
 Alignment Scores: Pred. No.: 0.397 Length: 75
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 45.00% Indels: 0
 DB: 22 Gaps: 0

US-10-060-830-1115 (1-60) x ABB42217 (1-75)
 QY 34 GATGGATGGACACACTGTACTAGGC 60
 Db 1 AspGlyCysGlyHisThrValLeuGly 9
 RESULT 7
 AAM63102
 ID AAM63102 standard; Protein; 75 AA.
 XX
 AC AAM63102;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 35207.
 XX
 KW Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157275-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00667.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-483446/52.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -

XX Example 4; SEQ ID NO: 35207; 650pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 XX
 XX Sequence 75 AA;
 SQ
 Alignment Scores: Pred. No.: 0.397 Length: 75
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 45.00% Indels: 0
 DB: 22 Gaps: 0

US-10-060-830-1115 (1-60) x AAM63102 (1-75)
 QY 34 GATGGATGGACACACTGTACTAGGC 60
 Db 1 AspGlyCysGlyHisThrValLeuGly 9
 RESULT 8
 AAM75913
 ID AAM75913 standard; Protein; 75 AA.
 XX
 AC AAM75913;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 36219.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.
 XX
 OS Homo sapiens.
 XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 XX Example 4; SEQ ID NO: 36219; 658pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.

```

XX
SQ Sequence 75 AA;

Alignment Scores:
Pred. No.: 0.397 Length: 75
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 45.00% Indels: 0
DB: 22 Gaps: 0

US-10-060-830-1115 (1-60) x AAM75913 (1-75)
QY 34 GATGATGTGGACACACTGTACTAGGC 60
Db 1 AspGlyCysGlyHisThrValLeuGly 9

RESULT 9
AAM36024
ID AAM36024 standard; Protein: 75 AA.
XX
AC AAM36024;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #10061 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-48897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 36293; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP:
XX see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.
XX
SQ Sequence 75 AA;

Alignment Scores:
Pred. No.: 0.397 Length: 75
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 45.00% Indels: 0
DB: 22 Gaps: 0

US-10-060-830-1115 (1-60) x AAM36024 (1-75)
QY 34 GATGATGTGGACACACTGTACTAGGC 60
Db 1 AspGlyCysGlyHisThrValLeuGly 9

RESULT 10
ABG45349
ID ABG45349 standard; Peptide; 75 AA.
XX
AC ABG45349;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human peptide encoded by genome-derived single exon probe SEQ ID 35014.
XX
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
XX
PR 26-MAY-2000; 2000US-207456P.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-234687P.
XX
PR 27-SEP-2000; 2000US-236359P.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples -
XX
PS Claim 27; SEQ ID No 35014; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of
XX probes; the novel set of probes which hybridise at high stringency to a
XX nucleic acid expressed in the human lung; measuring gene expression in a
XX sample derived from human lung, comprising (a) contacting the array with
XX a collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of
XX the array; identifying exons in a eukaryotic genome, comprising
XX (a) algorithmically predicting at least one exon from genomic sequences
XX of the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several

```

CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC hamangioleiomyomatosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 75 AA;

Alignment Scores:
 Pred. No.: 0.397 Length: 75
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 45.00% Indels: 0
 DB: 23 Gaps: 0

US-10-060-830-1115 (1-60) x ABG45349 (1-75)

QY 34 GATGGATCGACACTGTACTAGGC 60

Db 1 AspGlyCysGlyHisThrValLeuGly 9
 |||||

RESULT 11

AAE22720

ID AAE22720 standard; peptide: 20 AA.

XX AC AAE22720;

DT 09-AUG-2002 (first entry)

XX DE Human neuropilin-Hy2 signal peptide.

XX Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; prozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
 KW immunosuppressive.

XX OS Homo sapiens.

XX PN WO200222815-A1.

XX PD 21-MAR-2002.

XX PF 12-SEP-2001; 2001WO-US28488.

XX PR 11-SEP-2000; 2000US-0659671.

PR 06-SEP-2001; 2001US-317902P.

XX (HYSE-) HYSEQ INC.

XX Tang YT;

XX WPI; 2002-393966/42.

XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects

XX Disclosure: Page 130-131; 152pp; English.

XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy2
 CC signal peptide.

XX Sequence 20 AA;

Alignment Scores:

Pred. No.: 5.53 Length: 20
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 40.00% Indels: 0
 DB: 23 Gaps: 0

US-10-060-830-1115 (1-60) x AAE22720 (1-20)

QY 1 CTCTGCTCGAGCAGCTGGAGCC 24

Db 13 LeuLeuLeuGluAspAlaGlyAla 20
 |||||

RESULT 12

AAE44100

ID AAB44100 standard; Protein: 22 AA.

XX AC AAB44100;

XX DT 08-FEB-2001 (first entry)

XX DE Human cancer associated protein sequence SEQ ID NO:1545.

XX Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
 KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
 KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
 KW dermatological; neuroprotective; thrombolytic; coagulant; neurotropic;
 KW vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;

KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX Homo sapiens.
OS
XX WO200055350-A1.
PN
XX 21-SEP-2000.
PD
XX
XX 08-MAR-2000; 2000WO-US05882.
PF
XX 12-MAR-1999; 99US-0124270.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
PI
XX WPI: 2000-587533/55.
DR
XX N-PSDB; AAC78309.
DR
XX Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
PT
PS Claim 11; Page 2219; 2352pp; English.
XX
XX AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnarary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 22 AA;

Alignment Scores: 59.9 Length: 22
Pred. No.: 7.00 Matches: 7
Score: 7.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 21 Gaps: 0

US-10-060-830-1115 (1-60) x AAB44100 (1-22)

QY 45 TCCACATCCATCAGCTTGCTG 25
Db 7 SerThrSerIleThrLeuLeu 13

RESULT 13
ABG06004
ID ABG06004 standard; Protein; 36 AA.
XX
AC ABG06004;
XX
DT 13-FEB-2002 (first entry)
XX

DE Novel human diagnostic protein #5995.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR
XX N-PSDB; AAS70191.
DR
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PT
XX Claim 20; SEQ ID No 36363; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC ad and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 36 AA;

Alignment Scores: 54.9 Length: 36
Pred. No.: 7.00 Matches: 7
Score: 7.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 22 Gaps: 0

US-10-060-830-1115 (1-60) x ABG06004 (1-36)

QY 31 CTTGCTGGGCTCCAGGTCCT 11
Db 7 LeuAlaGlyLeuGlnArgPro 13

RESULT 14
AAB39047
ID AAB39047 standard; Protein; 59 AA.
XX
AC AAB39047;
XX

XX 02-FEB-2001 (first entry)
 DT Human secreted protein BLAST search protein SEQ ID NO: 137.
 DE
 XX Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
 KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 PN WO200056880-A1.
 XX
 XX 28-SEP-2000.
 PD
 XX 16-MAR-2000; 2000WO-US06781.
 PF
 XX 19-MAR-1999; 99US-0125363.
 PR
 XX 08-DEC-1999; 99US-0169617.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM, Komatsoulis G;
 PI
 XX WPI; 2000-602220/57.
 DR
 XX Nucleic acid molecules encoding human secreted proteins, used in
 PT preventing, treating or ameliorating disorders such as Parkinson's and
 PT Alzheimer's diseases, (cancers and infections -
 XX
 XX Disclosure; Page 401; 422pp; English.
 XX
 CC The invention relates to the isolation of genes AAC59679-c59728 encoding
 CC the human secreted proteins AAB38971-B39020. This sequence represents a
 CC peptide fragment homologous to the protein encoded by the gene isolated
 CC in the present invention. The sequence is a search result from a BLASTX
 CC homology search. The genes and proteins are useful for preventing,
 CC ameliorating or treating medical conditions, e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such
 CC as myocardial ischaemia; (d) wound healing; (e) neurological diseases
 CC e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such
 CC as viral, bacterial, fungal and parasitic infections.
 XX
 XX Sequence 59 AA;
 SQ
 Alignment Scores:
 Pred. No.: 50.2 Length: 59
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatve: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 21 Gaps: 0
 US-10-060-830-1115 (1-60) x AAB39047 (1-59)
 QY 45 TCCACATCCATCCTGCTGCTG 25
 Db 16 SerThrSerIleThrLeuLeu 22
 RESULT 15
 ID AAM06512
 XX AAM06512 standard; Protein; 76 AA.

AC AAM06512;
 XX 05-OCT-2001 (first entry)
 DT Human foetal protein, SEQ ID NO: 243.
 XX
 DE Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
 KW neurotropic; neuroprotective; thrombolytic; osteopathic; anti-inflammatory;
 KW gene therapy; antisenescence therapy; cancer; immune disorder;
 KW growth disorder; osteoporosis; thrombolytic disorder;
 KW nervous system disorder; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200155339-A2.
 XX
 XX 02-AUG-2001.
 PD
 XX 25-JAN-2001; 2001WO-US02723.
 PF
 XX 25-JAN-2000; 2000US-0491404.
 PR
 XX 15-SEP-2000; 2000US-0663870.
 PR
 XX 06-NOV-2000; 2000US-0707351.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
 PI Liu C, Asundi V, Zhou P, Werhman T;
 PI
 XX WPI; 2001-465571/50.
 DR
 XX N-PSDB; AAH94187.
 XX
 XX Novel fetal proteins useful for the treatment and diagnosis of diseases
 PT associated with dysfunction of the protein e.g. cancers, immune
 PT disorders, growth disorders, thrombolytic disorders, nervous system
 PT disorders and inflammation -
 XX
 XX Claim 10; Page 269; 715pp; English.
 XX
 CC The invention relates to novel foetal polypeptides encoded by
 CC polynucleotides comprising one of 477 sequences fully defined in the
 CC specification. The foetal polynucleotides and polypeptides are
 CC useful in the treatment and diagnosis of diseases such as cancers,
 CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
 CC disorders, nervous system disorders and inflammation. The present
 CC sequence is a polypeptide encoded by a cDNA assembled using
 CC an expressed sequence tag (EST) found to be expressed in human
 CC foetal tissue cDNA libraries.
 XX
 XX Sequence 76 AA;
 SQ
 Alignment Scores:
 Pred. No.: 48 Length: 76
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatve: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 22 Gaps: 0
 US-10-060-830-1115 (1-60) x AAM06512 (1-76)
 QY 38 CCATCAGCTTGTCTGGCTCCA 18
 Db 39 ProSerProCysTrpAlaPro 45
 Search completed: January 21, 2003, 14:49:04
 Job time : 7.4252 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:49:12 ; Search time 3.2126 Seconds
(without alignments)
3590.903 Million cell updates/sec

Title: US-10-060-830-1115

Perfect score: 20

Sequence: 1 ctgctgctcgaggacgctgg.....gtggacacactgtactagc 60

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565536

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US10060830/runat_16012003_092703_1513/app_query.fasta_1.1109
-DB=PIR_73 -QFMT=fastan -SUFFIX=olin2p.rpr -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.CGN_1.1.26 -runat_16012003_092703_1513 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMXP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	8	42.1	838	2 B38656	vacuolar proton pu
2	7	35.0	63	2 AC2278	hypothetical prote
3	7	35.0	107	2 T21842	hypothetical prote
4	7	35.0	107	2 T24886	hypothetical prote
C 5	7	36.8	143	2 T04524	photosystem I chal
6	7	35.0	227	2 H75506	oxidoreductase, sh
C 7	7	36.8	249	2 T46184	hypothetical prote
C 8	7	36.8	323	2 T07734	homeotic protein v
C 9	7	36.8	347	1 DNHUN2	NADH2 dehydrogenas
C 10	7	36.8	347	2 T11834	NADH2 dehydrogenas
C 11	7	36.8	347	2 B59153	NADH2 dehydrogenas
C 12	7	36.8	349	2 A44507	licheninase (EC 3.
C 13	7	36.8	368	2 G83463	probable methyltra
C 14	7	36.8	368	2 T46615	chemotaxis protein

C 15	7	36.8	431	2 T28697	hypothetical prote
C 16	7	36.8	500	1 JC4157	cytochrome P450 2D
C 17	7	35.0	543	2 A53310	phormone CAD1 bin
C 18	7	36.8	576	2 T46385	hypothetical prote
C 19	7	36.8	597	2 S00962	hypothetical prote
C 20	7	36.8	614	2 D84191	dolichol-P-glucose
C 21	7	36.8	621	1 S59632	endo-1,4-beta-xyla
C 22	7	36.8	649	2 D85135	hypothetical prote
C 23	7	36.8	767	2 S55618	hypothetical prote
C 24	7	36.8	1156	2 A47397	adducin homolog -
C 25	7	36.8	1237	2 A34598	ecdysone-induced p
C 26	7	36.8	1394	2 B34598	hypothetical prote
C 27	7	35.0	1535	2 T49042	hypothetical prote
C 28	7	36.8	1647	2 T49412	ribosomal protein
C 29	6	30.0	35	2 S32679	conserved hypothet
C 30	6	30.0	67	2 A10738	hypothetical prote
C 31	6	31.6	76	2 T03810	neutrophil cation
C 32	6	30.0	93	2 S14314	conserved hypothet
C 33	6	31.6	93	2 G87348	ymfJ protein - Esc
C 34	6	30.0	94	2 E64859	hypothetical prote
C 35	6	31.6	102	2 C96595	hypothetical prote
C 36	6	31.6	104	2 AF3201	hypothetical prote
C 37	6	30.0	107	2 T22709	hypothetical prote
C 38	6	30.0	107	2 T31471	hypothetical prote
C 39	6	30.0	107	2 T27502	hypothetical prote
C 40	6	30.0	107	2 T27713	hypothetical prote
C 41	6	30.0	113	2 G82387	hypothetical prote
C 42	6	31.6	117	2 T48736	hypothetical prote
C 43	6	30.0	117	2 T28562	hypothetical prote
C 44	6	30.0	117	2 B72166	Az1L protein - var
C 45	6	30.0	117	2 E42519	Az1L 13.6K protein

ALIGNMENTS

RESULT 1

B38656
vacuolar proton pump 116K chain - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-Feb-1992 #sequence_revision 21-Feb-1992 #text_change 22-Jun-1999
C:Accession: B38656
R:Perin, M.S.; Fried, V.A.; Stone, D.K.; Xie, X.S.; Suedhof, T.C.
J. Biol. Chem. 266, 3877-3881, 1991
A:Title: Structure of the 116-kDa polypeptide of the clathrin-coated vesicle/synaptic
A:Reference number: A38656; MUID:91139686; PMID:1704894
A:Accession: B38656
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-838 <PER>
A:Cross-references: GB:M58758; NID:g206429; PIDN:AAA41962.1; PID:g206430
C:Superfamily: vacuolar ATP synthase 95K chain
C:Keywords: alternative splicing; transmembrane protein

Alignment Scores:
Pred. No.: 2.02 Length: 838
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.11% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x B38656 (1-838)

Qy 33 ACCTGTGGGCTCCAGGCTCTC 10

|||||
Db 494 ThrLeuLeuGlySerSerValLeu 501

RESULT 2

AC2278
hypothetical protein asi3778 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C;Accession: AC2278
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
 A;Reference number: AB1807; MUID:21595285; PMID:11759840
 A;Accession: AC2278
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-63 <KOR>
 A;Cross-references: GB:BA000019; PIDN:BA075477.1; PID:gl7132912; GSPDB:GN00179
 A;Experimental source: strain PCC 7120
 C;Genetics:
 A;Gene: asl3778

Alignment Scores:
 Pred. No.: 34 Length: 63
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 35.00% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x AC2278 (1-63)
 Qy 17 CTGGAGCCCGCAGGTGATG 37
 Db 8 LeuGluProSerLysValMet 14
 |||||||
 |||||||

RESULT 3
 T21642
 Hypothetical protein F32B6.7 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 26-May-2000
 C;Accession: T21642
 R;Basham, V.
 Submitted to the EMBL Data Library, October 1996
 A;Reference number: Z19453
 A;Accession: T21642
 A;Status: preliminary; translated from GB/EMBL/DBD
 A;Molecule type: DNA
 A;Residues: 1-107 <WIL>
 A;Cross-references: EMBL:Z81074; PIDN:CAB03041.1; GSPDB:GN00022; CESP:F32B6.7
 A;Experimental source: clone F32B6
 C;Genetics:
 A;Gene: CESP:F32B6.7
 A;Map position: 4
 A;Introns: 33/3
 C;Superfamily: Caenorhabditis elegans hypothetical protein F32B6.7

Alignment Scores:
 Pred. No.: 31.2 Length: 107
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 35.00% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x T21642 (1-107)
 Qy 16 GCTGGAGCCCGCAGGTGAT 36
 Db 93 AlaglyAlagInGInGlyAsp 99
 |||||||
 |||||||

RESULT 4
 T24886
 Hypothetical protein T13F2.12 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 26-May-2000
 C;Accession: T24886
 R;Swinburne, J.
 Submitted to the EMBL Data Library, October 1996
 A;Reference number: Z19947

C;Accession: T24886
 A;Status: preliminary; translated from GB/EMBL/DBD
 A;Molecule type: DNA
 A;Residues: 1-107 <WIL>
 A;Cross-references: EMBL:Z81122; PIDN:CAB03363.1; GSPDB:GN00022; CESP:T13F2.12
 A;Experimental source: clone T13F2
 C;Genetics:
 A;Gene: CESP:T13F2.12
 A;Map position: 4
 A;Introns: 33/3
 C;Superfamily: Caenorhabditis elegans hypothetical protein F32B6.7

Alignment Scores:
 Pred. No.: 31.2 Length: 107
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 35.00% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x T24886 (1-107)
 Qy 16 GCTGGAGCCCGCAGGTGAT 36
 Db 93 AlaglyAlagInGInGlyAsp 99
 |||||||
 |||||||

RESULT 5
 T04524
 Photosystem I chain IV homolog F16A16.140 - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 21-May-1999
 C;Accession: T04524
 R;Bevan, M.; Brandt, P.; Dose, S.; Jarke, D.; Scharfe, M.; Schon, O.; Hoheisel, J.; M
 Submitted to the Protein Sequence Database, February 1999
 A;Reference number: Z15376
 A;Accession: T04524
 A;Molecule type: DNA
 A;Residues: 1-143 <BEV>
 A;Cross-references: EMBL:AL035353
 A;Experimental source: cultivar Columbia; BAC clone F16A16
 C;Genetics:
 A;Map position: 4
 A;Introns: 86/3; 108/3
 A;Note: F16A16.140
 C;Superfamily: photosystem I chain IV

Alignment Scores:
 Pred. No.: 29.8 Length: 143
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x T04524 (1-143)
 Qy 23 GCTCCAGCGTCTCGAGCAGC 3
 Db 50 AlaProAlaSerSerSerSer 56
 |||||||
 |||||||

RESULT 6
 H75506
 Oxidoreductase, short-chain dehydrogenase/reductase family - Deinococcus radiodurans
 C;Species: Deinococcus radiodurans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C;Accession: H75506
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: H75506

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-227 <WHI>
A:Cross-references: GB:AE001912; GB:AE000513; NID:g6458228; PIDN:AAF10118.1; PID:g645823
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0543
A:Map position: 1
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Alignment Scores:
Pred. No.: 27.7 Length: 227
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x H75506 (1-227)
QY 7 CTCGAGGCGTGGAGCCAG 27
|||||
Db 26 LeuGluaSPAlaGlyAlaGln 32

RESULT 7
T46184
hypothetical protein T8H10.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46184
R:Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Meves, H.W.; Lemcke, K.; Mayer, K.F.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23014
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <BEN>
A:Cross-references: EMBL:AL133248
A:Experimental source: cultivar Columbia; BAC clone T8H10
C:Genetics:
A:Map position: 3
A:Introns: 33/3
A:Note: T8H10.80

Alignment Scores:
Pred. No.: 27.3 Length: 249
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x T46184 (1-249)
QY 23 GCTCCAGCGTCTCGAGCAGC 3
|||||
Db 160 AlaProAlaSerSerSerSer 166

RESULT 8
T07734
homeotic protein VAHOX1 - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C:Accession: T07734
R:Tornero, P.; Conejero, V.; Vera, P.
Plant J. 9, 639-648, 1996
A:Title: Phloem-specific expression of a plant homeobox gene during secondary phases of
A:Reference number: Z16103; MUID:96237454; PMID:8653114
A:Accession: T07734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-323 <TOR>
A:Cross-references: EMBL:X94947; NID:g1161574; PIDN:CAA64417.1; PID:g1161575

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-227 <WHI>
A:Cross-references: GB:AE001912; GB:AE000513; NID:g6458228; PIDN:AAF10118.1; PID:g645823
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0543
A:Map position: 1
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Alignment Scores:
Pred. No.: 27.7 Length: 227
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x H75506 (1-227)
QY 7 CTCGAGGCGTGGAGCCAG 27
|||||
Db 26 LeuGluaSPAlaGlyAlaGln 32

RESULT 7
T46184
hypothetical protein T8H10.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46184
R:Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Meves, H.W.; Lemcke, K.; Mayer, K.F.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23014
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <BEN>
A:Cross-references: EMBL:AL133248
A:Experimental source: cultivar Columbia; BAC clone T8H10
C:Genetics:
A:Map position: 3
A:Introns: 33/3
A:Note: T8H10.80

Alignment Scores:
Pred. No.: 27.3 Length: 249
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x T46184 (1-249)
QY 23 GCTCCAGCGTCTCGAGCAGC 3
|||||
Db 160 AlaProAlaSerSerSerSer 166

RESULT 8
T07734
homeotic protein VAHOX1 - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C:Accession: T07734
R:Tornero, P.; Conejero, V.; Vera, P.
Plant J. 9, 639-648, 1996
A:Title: Phloem-specific expression of a plant homeobox gene during secondary phases of
A:Reference number: Z16103; MUID:96237454; PMID:8653114
A:Accession: T07734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-323 <TOR>
A:Cross-references: EMBL:X94947; NID:g1161574; PIDN:CAA64417.1; PID:g1161575

A:Experimental source: cultivar VFN8; leaf
C:Genetics:
A:Gene: VAHOX1
A:Introns: 49/3; 178/3
A:Note: specifically expressed in the phloem during phases of secondary growth
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:88-144/Domain: homeobox homology <HOX>

Alignment Scores:
Pred. No.: 26.2 Length: 323
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x T07734 (1-323)
QY 23 GCTCCAGCGTCTCGAGCAGC 3
|||||
Db 40 AlaProAlaSerSerSerSer 46

RESULT 9
DNHUN2
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - human mitochondrion
N:Alternate names: NADH-ubiquinone oxidoreductase chain 2
C:Species: mitochondrion Homo sapiens (man)
C:Date: 22-May-1981 #sequence_revision 17-Jul-1998 #text_change 03-Jun-2002
C:Accession: A0414; A38016; I52262; I65212
R:Anderson, S.; Bankier, A.T.; Barrell, B.G.; de Bruijn, M.H.L.; Coulson, A.R.; Droul
Nature 290, 457-465, 1981
A:Title: Sequence and organization of the human mitochondrial genome.
A:Reference number: A00151; MUID:81173052; PMID:7219534
A:Accession: A0414
A:Molecule type: DNA
A:Residues: 1-347 <AND>
A:Cross-references: GB:V00662; NID:gl3003; PIDN:CAA24027.1; PID:g578710; GSPDB:GN0010
A:Note: the authors translated the initiation codon ATT for residue 1 as Ile
R:Sanger, F.; Coulson, A.R.; Barrell, B.G.; Smith, A.J.H.; Roe, B.A.
J. Mol. Biol. 143, 161-178, 1980
A:Title: Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing.
A:Reference number: A38016; MUID:81170577; PMID:6260957
A:Accession: A38016
A:Molecule type: DNA
A:Residues: 37-347 <SAN>
A:Cross-references: GB:M10546; NID:g337297; PIDN:AAA55502.1; PID:g786121
A:Note: the authors do not show a translation for 1-36 and use 37-Met as the start co
R:Lin, F.H.; Lin, R.; Wisniewski, H.M.; Hwang, Y.W.; Grundke-Iqbal, I.; Healy-Louie,
Biochem. Biophys. Res. Commun. 182, 238-246, 1992
A:Title: Detection of point mutations in codon 331 of mitochondrial NADH dehydrogenas
A:Reference number: I52262; MUID:92118019; PMID:1370613
A:Accession: I52262
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 269-330, 'T', 332-347 <LIN>
A:Cross-references: GB:S75895; NID:g242496; PIDN:AAB20905.1; PID:g242497
A:Experimental source: mutation, Alzheimer's disease, clone 82-2
A:Accession: I65212
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 269-270, 278-330, 'S', 332, 'LDSVFLWLT', 342 <LIN>
A:Cross-references: GB:S75896; NID:g242498; PIDN:AAB20906.1; PID:g242499
C:Genetics:
A:Gene: GDB:MTN2
A:Cross-references: GDB:118912; OMIM:516001
A:Map position: MTH4470-5511
A:Genome: mitochondrion
A:Genetic code: SGC1
A:Start codon: ATT
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Alignment Scores:

Pred. No.: 25.9 Length: 347
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x DNHUN2 (1-347)

QY 45 TCACATCCATCACCCTTGCTG 25

Db 299 SerThrSerIleThrLeuLeu 305

RESULT 10

Tl1834

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - common gibbon mitochondrion
 C:Species: mitochondrion Hylobates lar (common gibbon, white-handed gibbon)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002

C:Accession: Tl1834

R:Aranson, U.; Gullberg, A.; Xu, X.

Hereditas 124, 185-189, 1996

A:Title: A complete mitochondrial DNA molecule of the white-handed gibbon, Hylobates lar

A:Reference number: Z17353

A:Accession: Tl1834

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-347 <ARN>

A:Cross-references: EMBL:X99256; PIDN:CAA67629.1

A:Experimental source: isolate Ester

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SGC1

A:Note: NADH2

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Alignment Scores:

Pred. No.: 25.9 Length: 347
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x Tl1834 (1-347)

QY 45 TCACATCCATCACCCTTGCTG 25

Db 299 SerThrSerIleThrLeuLeu 305

RESULT 11

B59153

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - western lowland gorilla mitochondrion

A:Alternate names: NADH-ubiquinone oxidoreductase chain 2

C:Species: mitochondrion Gorilla gorilla gorilla (western lowland gorilla)

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Jun-2002

C:Accession: B59153

R:Xu, X.; Aranson, U.

Mol. Biol. Evol. 13, 691-698, 1996

A:Title: A complete sequence of the mitochondrial genome of the western lowland gorilla.

A:Reference number: Z17269; MUID:96212991; PMID:8676744

A:Accession: B59153

A:Status: preliminary; nucleic acid sequence not shown; translation not shown; translated

A:Molecule type: DNA

A:Residues: 1-347 <XUX>

A:Cross-references: GB:X93347; NID:91304307; GSPDB:GN00106

A:Note: submitted to GenBank, November 1995

C:Genetics: this translation is not annotated in GenBank entry GCMITG, release 114.0

A:Gene: NADH2

A:Genome: mitochondrion

A:Genetic code: SGC1

A:Start codon: ATT

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Alignment Scores:

Pred. No.: 25.9 Length: 347
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x B59153 (1-347)

QY 45 TCACATCCATCACCCTTGCTG 25

Db 299 SerThrSerIleThrLeuLeu 305

RESULT 12

A44507

licheninase (EC 3.2.1.73) - Fibrobacter succinogenes

C:Species: Fibrobacter succinogenes

C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999

C:Accession: A44507

R:Teather, R.M.; Erfle, J.D.

J. Bacteriol. 172, 3837-3841, 1990

A:Title: DNA sequence of a Fibrobacter succinogenes mixed-linkage beta-glucanase (1,3

A:Reference number: A44507; MUID:90299807; PMID:2193918

A:Accession: A44507

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <TEA>

A:Cross-references: EMBL:M33676; NID:9148575; PIDN:AAA24896.1; PID:9148576

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Alignment Scores:

Pred. No.: 25.9 Length: 349
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x A44507 (1-349)

QY 23 GCTCCAGCGTCTCGAGCAGC 3

Db 277 AlaProAlaSerSerSerSer 283

RESULT 13

G83463

probable methyltransferase PA1459 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G83463

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: G83463

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-368 <STO>

A:Cross-references: GB:AE004575; GB:AE004091; NID:99947404; PIDN:AAG04848.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1459

C:Superfamily: protein-glutamate methyltransferase; response regulator homology

Alignment Scores:

Pred. No.: 25.7 Length: 368
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x G83463 (1-368)

QY 23 GCTCCAGGCTCCTCGAGCAGC 3

Db 151 AlaProAlaSerSerSer 157

RESULT 14

T46615

chemotaxis protein cheB [imported] - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa

C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 04-Mar-2000

C:Accession: T46615

R:Kato, J.; Nakamura, T.; Kuroda, A.; Ohtake, H.

submitted to the EMBL Data Library, April 1998

A:Description: Cloning, sequence and characterization of chemotaxis genes in Pseudomonas

A:Reference number: Z23079

A:Accession: T46615

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-368 <KAT>

A:Cross-references: EMBL:AB012767; PIDN:BAA33550.1

A:Experimental source: strain PAO1

C:Genetics:

A:Note: cheB

C:Superfamily: protein-glutamate methylesterase; response regulator homology

Alignment Scores:

Pred. No.: 25.7 Length: 368
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x T46615 (1-368)

QY 23 GCTCCAGGCTCCTCGAGCAGC 3

Db 151 AlaProAlaSerSerSer 157

RESULT 15

T28697

hypothetical protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T28697

R:Parkhill, J.; Bentley, S.D.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, January 1999

A:Reference number: Z20512

A:Accession: T28697

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-431 <PAR>

A:Cross-references: EMBL:AL023496; PIDN:CAA18914.1

C:Superfamily: tetracycline resistance protein

Alignment Scores:

Pred. No.: 25 Length: 431
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x T28697 (1-431)

QY 31 CTTGCTGGGCTCCAGGTCCT 11

Db 125 LeuAlaGlyLeuGlnArgPro 131

Search completed: January 21, 2003, 14:54:48
Job time : 6.2126 secs

DR EMBL; M58758; AAA41962.1; --
DR PIR; B38656; B38656.
DR InterPro; IPR002490; V_Atpase_sub116.
DR Pfam; PF01496; V_Atpase_sub_a; 1.
KW Hydrogen ion transport; Transmembrane; Glycoprotein;
KW Alternative splicing.
FT DOMAIN 1 395 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 396 425 POTENTIAL.
FT TRANSMEM 426 465 POTENTIAL.
FT TRANSMEM 466 563 POTENTIAL.
FT TRANSMEM 564 593 POTENTIAL.
FT TRANSMEM 594 634 POTENTIAL.
FT TRANSMEM 635 654 POTENTIAL.
FT TRANSMEM 655 794 POTENTIAL.
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 706 711 MISSING (IN ISOFORM II).
SQ SEQUENCE 838 AA; 96327 MW; D25FAD918638D7EE CRC64;

Alignment Scores:
Pred. No.: 0.904 Length: 838
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.11% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x VPPL_RAT (1-838)
QY 33 ACCTTGGTGGCTCCAGGCTCCTC 10
Db 494 ThrLeuLeuGlySerSerValleu 501

RESULT 2
VPPL_MOUSE
ID VPPL_MOUSE STANDARD; PRT; 839 AA.
AC Q9JL14; Q9JL14; Q9JHJ4;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vacuolar proton translocating ATPase 116 kDa subunit A isoform 1
DE (Clathrin-coated vesicle/synaptic vesicle proton pump 116 kDa subunit)
DE (Vacuolar proton pump subunit 1) (Vacuolar adenosine triphosphatase subunit Acl16).
DE ATP6V0A1 OR ATP6N1.
OS Mus musculus (Mouse).
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM AI-II).
RA Howell M.L., Dean G.E.;
RT "CDNA sequences for mouse vacuolar ATPase subunits.";
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Liver;
RX MEDLINE=20167151; PubMed=10702241;
RA Nishi T., Forgac M.;
RT "Molecular cloning and expression of three isoforms of the 100-kDa a subunit of the mouse vacuolar proton-translocating ATPase.";
RL J. Biol. Chem. 275:6824-6830(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM AI-III).
RX MEDLINE=20187595; PubMed=10722719;
RA Toyomura T., Oka T., Yamaguchi C., Wada Y., Futai M.;
RT "Three subunit A isoforms of mouse vacuolar H⁺-ATPase. Preferential expression of the a3 isoform during osteoclast differentiation.";
RL J. Biol. Chem. 275:8760-8765(2000).
CC -1- FUNCTION: REQUIRED FOR ASSEMBLY AND ACTIVITY OF THE VACUOLAR ATPASE. POTENTIAL ROLE IN DIFFERENTIAL TARGETING AND REGULATION OF THE ENZYME FOR A SPECIFIC ORGANELLE (BY SIMILARITY).
CC -1- SUBUNIT: THE VACUOLAR PROTON PUMP FROM CLATHRIN-COATED VESICLES

CC CONTAINS EIGHT TO NINE DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. COATED VESICLE.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; AI-I, AI-II (SHOWN HERE) AND AI-III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE V-ATPASE 116 kDa SUBUNIT FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC
CC EMBL; UI3836; AAC83083.1; --
CC EMBL; AF218249; AAF59918.1; --
CC EMBL; AF218250; AAF59919.1; --
CC EMBL; AF218251; AAF59920.1; --
CC EMBL; AF218252; AAF59921.1; --
CC EMBL; AB022321; BAA93005.1; --
CC MGD; MGI:103286; Atp6n1.
CC InterPro; IPR002490; V_Atpase_sub116.
CC Pfam; PF01496; V_Atpase_sub_a; 1.
KW Hydrogen ion transport; Transmembrane; Glycoprotein;
KW Alternative splicing; Polymorphism.
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 496 496 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 142 148 MISSING (IN ISOFORM AI-I AND ISOFORM AI-III).
FT VARSPLIC 712 712 E -> EPTDEV (IN ISOFORM AI-I).
FT VARIANT 194 194 F -> L.
FT CONFLICT 36 36 Q -> N (IN REF. 1).
FT CONFLICT 88 92 VPFP -> APLPW (IN REF. 1);
FT CONFLICT 112 112 N -> D (IN REF. 1).
FT CONFLICT 190 190 F -> S (IN REF. 1).
FT CONFLICT 262 262 K -> T (IN REF. 1).
FT CONFLICT 337 337 L -> P (IN REF. 1).
FT CONFLICT 357 357 S -> F (IN REF. 1).
FT CONFLICT 415 415 G -> R (IN REF. 1).
FT CONFLICT 518 518 F -> L (IN REF. 1).
FT CONFLICT 683 683 G -> W (IN REF. 1).
FT CONFLICT 803 803 H -> L (IN REF. 1).
SQ SEQUENCE 839 AA; 96500 MW; B3400474C8666B6C CRC64;

Alignment Scores:
Pred. No.: 0.904 Length: 839
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.11% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x VPPL_MOUSE (1-839)
QY 33 ACCTTGGTGGCTCCAGGCTCCTC 10
Db 501 ThrLeuLeuGlySerValleu 508

RESULT 3
NIC1_HUMAN
ID NIC1_HUMAN STANDARD; PRT; 99 AA.
AC Q9UGL9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE NICE-1 protein.
GN NICE1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

RC TISSUP=Keratinocytes;
RX MEDLINE=21154910; PubMed=11230159;
RA Marenholz I., Zirra M., Fischer D.F., Backendorf C., Ziegler A.,
RA Mischke D.;
RT "Identification of human epidermal differentiation complex (EDC)-
RT encoded genes by subtractive hybridization of entire YACs to a gridged
RT keratinocyte cDNA library.";
RL Genome Res. 11:341-355(2001);
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ243662; CAB65093.1; POLY-SER.
FT DOMAIN 35 38
FT DOMAIN 65 71
FT DOMAIN 75 78
FT POLY-CYS.
SQ SEQUENCE 99 AA; 9736 MW; B522D8E206305FDB CRC64;
Alignment Scores:
Pred. No.: 13.9 Length: 99
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 1 Gaps: 0
US-10-060-830-1115 (1-60) x NIC1_HUMAN (1-99)
QY 23 GCTCCAGCGCTCCTCGAGCAGC 3
Db 32 AlaProAlaSerSerSer 38
RESULT 4
PSEL_ARATH
ID PSEL_ARATH STANDARD; PRT; 143 AA.
AC Q9S831;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem I reaction center subunit IV A, chloroplast precursor
DE (PST-E A)
GN PSAB1 OR AF4G28750 OR F16A16.140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Legen J., Misera S., Herrmann R.G., Altschmid L.;
RT "Sequences and map position of 31 Arabidopsis thaliana cDNAs encoding
RT organellar polypeptides.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
RA Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidtheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohnel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,

```

```

RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzner E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buysnaert C., Gielens J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLaughlin K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Gabel S., de Haan M., Marse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel S., Fuchs M., Farmann B., Grandrath K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Massen O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefod F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
RA Perez-Perez A., Fumelle B., Bent E., Johnson S., Tacon D., Jesse C.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bleile C.,
RA Frishman D., Haase D., Lemcke K., Meves H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parknell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Hannon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spleth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:769-777(1999).
CC -!- FUNCTION: May form complexes with ferredoxin and ferredoxin-
CC oxidoreductase in photosystem I (PS I) reaction center (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane (By
CC similarity).
CC -!- PTM: 2 isoforms may exist. With or without the N-terminal alanine
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PSAE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ245908; CAB52678.1; -
DR EMBL; AL035353; CAA22977.1; -
DR EMBL; AL161573; CAB81463.1; -
DR HSP; O9WNP1.1Q2.
DR InterPro: IPR003375; PSI_Psae.
DR Pfam: PF02427; PSI_Psae_1.
KW Photosynthesis; Photosystem I; Chloroplast; Transist peptide;
KW Thylakoid; Membrane; Multigene family.
FT TRANSIT 1 44
FT CHAIN 45 143
FT MOD_RES 45 45 REMOVED POST-TRANSLATIONALLY (BY
FT SIMILARITY).
SQ SEQUENCE 143 AA; 14967 MW; 0E055D8B2F1D03F0 CRC64;
Alignment Scores:
Pred. No.: 13.2 Length: 143
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 36.84% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x PSEL_ARATH (1-143)

Oy 23 GCTCCAGCGTCTCGAGCAGC 3
|||||
Db 50 AlaProAlaSerSerSer 56

RESULT 5

HXDD_HUMAN

ID HXDD_HUMAN STANDARD: PRT; 335 AA.

AC P35453;

DT 01-JUN-1994 (Rel. 29, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Homeobox protein Hox-D13 (Hox-41).

GN HOXD13 OR HOX41.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96194671; PubMed=8614804;

RA Muragaki Y., Mundlos S., Upton J., Olsen B.R.;

RT "Altered growth and branching patterns in synpolydactyly caused by

mutations in HOXD13.";

RL Science 272:548-551(1996).

RN [2]

RP SEQUENCE FROM N.A.

RA Arai Y., Arai K., Kita K., Miwa H., Kamada N., Ohki M.;

RT "The t(2;11)(q31;p15) translocation in acute myeloid leukemia fuses

NU98 nucleoporin gene to HOXD13 homeobox gene.";

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 268-333 FROM N.A.

RX MEDLINE=91257849; PubMed=1675198;

RA D'Esposito M., Morelli F., Acampora D., Migliaccio E., Simeone A.,

RA Boncinelli E.;

RT "EVX2, a human homeobox gene homologous to the even-skipped

segmentation gene, is localized at the 5' end of HOX4 locus on

chromosome 2.";

RL Genomics 10:43-50(1991).

RN [4]

RP VARIANT SYNPOLYDACTYLY POLY-ALA INS.

RX MEDLINE=96414300; PubMed=8817328;

RA Akarsu A.N., Stollow I., Yilmaz E., Sayli B.S., Sarfarazi M.;

RT "Genomic structure of HOXD13 gene: a nine polyalanine duplication

causes synpolydactyly in two unrelated families.";

RL Hum. Mol. Genet. 5:945-952(1996).

CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF

A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH

SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- DISEASE: DEFECTS IN HOXD13 ARE A CAUSE OF SYNPOLYDACTYLY, ALSO

KNOWN AS SYNDACTYLY, TYPE II.

CC -1- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-sib.ch).

CC -----

DR EMBL; AF005220; AAC51635.1;

DR EMBL; AF005219; AAC51635.1; JOINED.

DR EMBL; AB032481; BAA95352.1;

DR PIR; B39065; B39065.

DR HSSP; P14653; 1B72.

DR TRANSFAC; T03335;

DR Genew; HGNC:5136; HOXD13.
DR MIM; 142989;
DR MIM; 186000;
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation; Disease mutation.
FT DOMAIN 17..22 POLY-SER.
FT DOMAIN 24..28 POLY-ALA.
FT DOMAIN 49..63 POLY-ALA.
FT DOMAIN 77..84 POLY-SER.
FT DOMAIN 105..110 POLY-ALA.
FT DNA_BIND 268..327 HOMEBOX.
FT VARIANT 49..49 A -> AAAAAAAAAA (IN SYNPOLYDACTYLY).
FT /FTID=VAR_003818.
SQ SEQUENCE 335 AA; 35210 MW; 0558D7B29F9B6E3E CRC64;

Alignment Scores:
Pred. No.: 11.6 Length: 335
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x HXDD_HUMAN (1-335)

Oy 23 GCTCCAGCGTCTCGAGCAGC 3

|||||

Db 14 AlaProAlaSerSerSer 20

RESULT 6

HXDD_MOUSE

ID HXDD_MOUSE STANDARD; PRT; 339 AA.

AC P70217; Q64177;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Homeobox protein Hox-D13 (Hox-4.8).

GN HOXD13 OR HOX-4.8.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97133301; PubMed=8978698;

RA Herault Y., Hraba-Reveney S., van der Hoeven F., Duboule D.;

RT "Function of the Evx-2 gene in the morphogenesis of vertebrate

limbs.";

RL EMBL J. 15:6727-6738(1996).

RN [2]

RP SEQUENCE OF 258-339 FROM N.A.

RX MEDLINE=92144412; PubMed=1685889;

RA Dalle P., Ippisua-Belmonte J.C., Boncinelli E., Duboule D.;

RT "The Hox-4.8 gene is localized at the 5' extremity of the Hox-4

complex and is expressed in the most posterior parts of the body

during development.";

RL Mech. Dev. 36:3-13(1991).

CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF

A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH

SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X99291; CAA67675.1; -;
 DR EMBL; S80553; AAB21366.2; -;
 DR HSSP; P14653; 1B72.
 DR MGD; MGI:96205; Hoxd13.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00071; HOMEBOX_2; 1.
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation.
 FT DOMAIN 25 29 POLY-SER.
 FT DOMAIN 53 67 POLY-ALA.
 FT DOMAIN 81 88 POLY-SER.
 FT DOMAIN 105 114 POLY-ALA.
 FT DNA_BIND 272 331 HOMEBOX.
 FT CONFLICT 304 307 R -> T (IN REF. 2).
 FT CONFLICT 307 307 A -> R (IN REF. 2).
 SQ SEQUENCE 339 AA; 35903 MW; 84AD2DDF4A56E498 CRC64;

Alignment Scores: 11.6 Length: 339
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36/84% Indels: 0
 DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x HXDD_MOUSE (1-339)

QY 23 GTCACGCTCCTCGAGCAGC 3

Db 22 AlaProAlaSerSerSerSer 28

RESULT 7

NU2M_HUMAN STANDARD; PRT; 347 AA.
 ID NU2M_HUMAN STANDARD; PRT; 347 AA.
 AC P03891; Q34769; Q9TGI0; Q9TGI1; Q9TGI2; Q9TGI3; Q9TGI4;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
 GN MTND2 OR ND2.
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-81173052; PubMed-7219534;
 RA Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L.,
 RA Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A.,
 RA Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G.;
 RT "Sequence and organization of the human mitochondrial genome.";
 RL Nature 290:457-465(1981).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-81170577; PubMed-6260957;
 RA Sanger F., Coulson A.R., Barrell B.G., Smith A.J.H., Roe B.A.;
 RT "Cloning in single-stranded bacteriophage as an aid to rapid DNA
 RT sequencing";
 RL J. Mol. Biol. 143:161-178(1980).
 RN [3]
 RP SEQUENCE FROM N.A., AND VARIANTS V-69; S-88; D-150; M-237; T-265;
 RP T-278; V-265 AND A-333.
 RX MEDLINE-98133898; PubMed-9475751;
 RA Wise C.A., Srna M., Eastaugh S.;

RT "Departure from neutrality at the mitochondrial NADH dehydrogenase
 RT subunit 2 gene in humans, but not in chimpanzees.";
 RL Genetics 148:409-421(1998).
 RN [4]
 RP SEQUENCE FROM N.A., AND VARIANTS ILE-43; LEU-325 AND THR-331.
 RC TISSUE=Placenta;
 RX MEDLINE-95132634; PubMed-7530363;
 RA Horai S., Hayasaka K., Kondo R., Tsugane K., Takahata N.;
 RT "Recent African origin of modern humans revealed by complete sequences
 RT of hominoid mitochondrial DNAs";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:532-536(1995).
 RN [5]
 RP IDENTIFICATION OF PROTEIN.
 RX MEDLINE-85188293; PubMed-3921850;
 RA Chomyn A., Mariottini P., Cleeter M.W.J., Ragan C.I., Matsuno-Yagi A.,
 RA Hatefi Y., Doolittle R.F., Attardi G.;
 RT "Six unidentified reading frames of human mitochondrial DNA encode
 RT components of the respiratory-chain NADH dehydrogenase.";
 RL Nature 314:592-597(1985).
 RN [6]
 RP VARIANT LHON ASP-150.
 RX MEDLINE-91144615; PubMed-19000003;
 RA Johns D.R., Berman J.;
 RT "Alternative, simultaneous complex I mitochondrial DNA mutations in
 RT Leber's hereditary optic neuropathy.";
 RL Biochem. Biophys. Res. Commun. 174:1324-1330(1991).
 RN [7]
 RP VARIANTS L-42; R-63; A-119; P-148; S-150; T-159 AND A-185.
 RX MEDLINE-92098084; PubMed-1757091;
 RA Marzuki S., Noer A.S., Lertrit P., Thyagarajan D., Kapsea R.,
 RA Utthanaphol P., Byrne E.;
 RT "Normal variants of human mitochondrial DNA and translation products:
 RT the building of a reference data base.";
 RL Hum. Genet. 88:139-145(1991).
 RN [8]
 RP VARIANT LHON SER-259.
 RX MEDLINE-92120513; PubMed-1732158;
 RA Brown M.D., Voljavec A.S., Lott M.T., Torroni A., Yang C.C.,
 RA Wallace D.C.;
 RT "Mitochondrial DNA complex I and III mutations associated with
 RT Leber's hereditary optic neuropathy.";
 RL Genetics 130:163-173(1992).
 RN [9]
 RP VARIANT AD SER-331.
 RX MEDLINE-92118019; PubMed-1370613;
 RA Lin F.-H., Lin R., Wisniewski H.M., Hwang Y.-W., Grundke-Iqbal I.,
 RA Healy-Louie G., Iqbal K.;
 RT "Detection of point mutations in codon 331 of mitochondrial NADH
 RT dehydrogenase subunit 2 in Alzheimer's brains.";
 RL Biochem. Biophys. Res. Commun. 182:238-246(1992).
 RN [10]
 RP VARIANT THR-57.
 RX MEDLINE-98127994; PubMed-9461455;
 RA Rieder M.J., Taylor S.L., Tobe V.O., Nickerson D.A.;
 RT "Automating the identification of DNA variations using quality-based
 RT fluorescence re-sequencing: analysis of the human mitochondrial
 RT genome.";
 RL Nucleic Acids Res. 26:967-973(1998).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane.
 CC -1- DISEASE: DEFECTS IN MTND2 ARE ONE OF THE CAUSES OF LEBER'S
 CC HEREDITARY OPTIC NEUROPATHY (LHON), A MATERNALLY INHERITED DISEASE
 CC RESULTING IN ACUTE BILATERAL BLINDNESS DUE TO RETINAL DEGENERATION
 CC PREDOMINANTLY IN YOUNG MEN. CARDIAC CONDUCTION DEFECTS AND
 CC NEUROLOGICAL DEFECTS HAVE ALSO BEEN DESCRIBED. RESULTING IN OPTIC
 CC NERVE DEGENERATION AND CARDIAC DYSRHYTHMIA.
 CC -1- DISEASE: DEFECTS IN MTND2 COULD BE ASSOCIATED WITH ALZHEIMER'S
 CC DISEASE (AD).
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL; J01415; AAB58944.1; -
CC EMBL; V00662; CAA24027.1; -
CC EMBL; M10546; AAG65502.1; ALT_INIT.
CC EMBL; D38112; BAA07291.1; -
CC EMBL; AF014882; AAC25441.1; -
CC EMBL; AF014884; AAC25443.1; -
CC EMBL; AF014885; AAC25444.1; -
CC EMBL; AF014887; AAC25446.1; -
CC EMBL; AF014889; AAC25448.1; -
CC EMBL; AF014890; AAC25449.1; -
CC EMBL; AF014891; AAC25450.1; -
CC EMBL; AF014892; AAC25451.1; -
CC EMBL; AF014895; AAC25454.1; -
CC EMBL; AF014896; AAC25455.2; -
CC EMBL; AF014897; AAC25456.1; -
CC EMBL; AF014898; AAC25457.1; -
CC EMBL; AF014899; AAC25458.2; -
CC EMBL; AF014900; AAC25459.1; -
CC EMBL; AF014901; AAC25460.1; -
CC PIR; A00414; DNHUN2.
CC Genew; HGNC:7456; MTND2.
CC MIM; 502500; -
CC MIM; 516001; -
CC MIM; 535000; -
CC InterPro; IPR003917; NADHub_oxred2.
CC InterPro; IPR001750; Oxidored_q1.
CC Pfam; PF00361; oxidored_q1.1.
CC PRINTS; PR01436; NADHDHGNASE2.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane;
CC Disease mutation; Leber's hereditary optic neuropathy;
CC Alzheimer's disease; Polymorphism.
FT VARIANT 42 P -> L.
FT VARIANT 43 V -> I.
FT VARIANT 57 I -> T.
FT VARIANT 63 Q -> R.
FT VARIANT 69 I -> V.
FT VARIANT 88 N -> S.
FT VARIANT 119 T -> A.
FT VARIANT 148 S -> P.
FT VARIANT 150 N -> D (IN LHON; SECONDARY MUTATION; DOES NOT SEEM TO DIRECTLY CAUSE THE DISEASE).
FT VARIANT 150 N -> S.
FT VARIANT 159 I -> T.
FT VARIANT 185 T -> A.
FT VARIANT 237 L -> M.
FT VARIANT 259 G -> S (IN LHON; RARE PRIMARY MUTATION).
FT VARIANT 265 A -> T.
FT VARIANT 265 A -> V.
FT VARIANT 278 I -> T.
FT VARIANT 325 F -> L.

```

```

FT VARIANT 331 331 /FTId=VAR_011355.
FT A -> S (IN AD).
FT /FTId=VAR_004758.
FT VARIANT 331 331 A -> T.
FT /FTId=VAR_004757.
FT T -> A.
FT /FTId=VAR_011356.
FT SQ SEQUENCE 347 AA; 38961 MW; C06FD982317C3F2D CRC64;

Alignment Scores:
Pred. No.: 11.5 Length: 347
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x NU2M_HUMAN (1-347)

QY 45 TCCACATCCATCACCTTGCTG 25
|||||
Db 299 SerThrSerIleThrLeuLeu 305

RESULT 8
NU2M_HYLLA
ID NU2M_HYLLA STANDARD; PRT; 347 AA.
AC Q95704;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 13-JUL-1997 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
GN MTND2 OR ND2 OR NADH2.
OS Hylobates lar (Common gibbon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.
OX NCBI_TaxID=9580;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ester;
RA Arnason U., Gullberg A., Xu X.;
RT "A complete mitochondrial DNA molecule of the white-handed gibbon, of
RT Hylobates lar, and comparison among individual mitochondrial genes of
RT all hominoid genera.";
RL Hereditas 124:185-189(1996).
CC -I- CATALYTIC ACTIVITY: NADH + ubiquinone -> NAD(+) + ubiquinol.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC -I- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

EMBL; X99256; CAA67629.1; -
DR InterPro; IPR003917; NADHub_oxred2.
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1.1.
DR PRINTS; PR01436; NADHDHGNASE2.
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 347 AA; 38736 MW; 0060C7BAFDCEFD28 CRC64;

Alignment Scores:
Pred. No.: 11.5 Length: 347
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 1 Gaps: 0

```

US-10-060-830-1115 (1-60) x NU2M_HYLLA (1-347)

QY 45 TCACATCCATCACCCTGCTG 25
|||||
Db 299 SerThrSerIleThrLeuLeu 305

RESULT 9

NU2M_PANTR
ID NU2M_PANTR STANDARD; PRT; 347 AA.
AC Q21798;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
GN MTND2 OR ND2.
OS Pan troglodytes (Chimpanzee).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98133898; PubMed=9475751;
RA Wise C.A., Srami M., Eastel S.;
RT "Departure from neutrality at the mitochondrial NADH dehydrogenase
subunit 2 gene in humans, but not in chimpanzees.";
RL Genetics 148:409-421(1998)
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
inner membrane.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF014908; AAC25467.1; -
DR EMBL; AF014909; AAC25468.1; -
DR EMBL; AF014910; AAC25469.1; -
DR EMBL; AF014911; AAC25470.1; -
DR EMBL; AF014912; AAC25471.1; -
DR EMBL; AF014913; AAC25472.1; -
DR EMBL; AF014914; AAC25473.1; -
DR EMBL; AF014915; AAC25474.1; -
DR EMBL; AF014916; AAC25475.1; -
DR EMBL; AF014917; AAC25476.1; -
DR EMBL; AF014918; AAC25477.1; -
DR EMBL; AF014919; AAC25478.1; -
DR EMBL; AF014920; AAC25479.1; -
DR EMBL; AF014921; AAC25480.1; -
DR InterPro: IPR003917; NADHUb_Oxred2.
DR InterPro: IPR001750; Oxidored_q1.
DR Pfam: PF00361; oxidored_q1; 1.
DR PRINTS; PR01436; NADHDHGNASE2.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
FT VARIANT 94 P -> S (IN STRAIN A-292).
SQ SEQUENCE 347 AA; 39020 MW; 2E8269D105810DAE CRC64;

Alignment Scores:
Pred. No.: 11.5 Length: 347
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x NU2M_PANTR (1-347)

QY 45 TCACATCCATCACCCTGCTG 25
|||||
Db 299 SerThrSerIleThrLeuLeu 305

RESULT 10

GUB_FIBSU
ID GUB_FIBSU STANDARD; PRT; 349 AA.
AC P17989;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Beta-glucanase precursor (EC 3.2.1.73) (Endo-beta-1,3-1,4 glucanase)
DE (1,3-1,4-beta-D-glucan 4-glucanohydrolase) (Mixed linkage beta-
glucanase) (Lichenase).
OS Fibrobacter succinogenes (Bacteroides succinogenes).
OC Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;
OC Fibrobacter.
OX NCBI_TaxID=833;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-57.
RC STRAIN=Isolate S85;
RX MEDLINE=90299807; PubMed=2193918;
RA Teather R.M., Erfile J.D.;
RT "DNA sequence of a Fibrobacter succinogenes mixed-linkage
beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene.";
RL J. Bacteriol. 172:3837-3841(1990).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glycosidic linkages
in beta-D-glucans containing 1,3- and 1,4-bonds.
CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M33676; AAA24896.1; -
DR PIR; A44507; A44507.
DR HSP; P23904; IAJK.
DR InterPro: IPR000757; Glyco_hydro_16.
DR Pfam; PF00722; Glyco_hydro_16; 1.
DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
KW Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 27
FT CHAIN 28 349 BETA-GLUCANASE.
FT ACT_SITE 79 79 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 83 83 PROTON DONOR (BY SIMILARITY).
FT DOMAIN 271 307 5 X 7 AA TANDEM REPEATS OF P-X-S-S-S-X.
FT REPEAT 271 277 1.
FT REPEAT 278 284 2.
FT REPEAT 285 291 3.
FT REPEAT 292 298 4.
FT REPEAT 301 307 5.
SQ SEQUENCE 349 AA; 37737 MW; 16DC4F5BDFC578A CRC64;

Alignment Scores:
Pred. No.: 11.5 Length: 349
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x GUB_FIBSU (1-349)

QY 23 GCTCAGCGTCTCGAGCAGC 3
|||||
Db 277 AlaProAlaSerSerSerSer 283

RESULT 11

CHEB_PSEAE

ID AC 087125; STANDARD; PRT; 368 AA.
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE protein-glutamate methyltransferase (EC 3.1.1.61).
 GN CHEB OR PA1459.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=99161286; PubMed=10052136;
 RA Kato J., Nakamura T., Kuroda A., Ohtake H.;
 RT "Cloning and characterization of chemotaxis genes in Pseudomonas
 aeruginosa";
 RL Biosci. Biotechnol. Biochem. 63:155-161(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen";
 RL Nature 406:959-964(2000).
 CC -1- FUNCTION: INVOLVED IN THE MODULATION OF THE CHEMOTAXIS SYSTEM.
 CC CATALYZES THE DEMETHYLATION OF SPECIFIC METHYLGLUTAMATE RESIDUES
 CC INTRODUCED INTO THE CHEMORECEPTORS (METHYL-ACCEPTING CHEMOTAXIS
 CC PROTEINS) BY CHEB (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Protein L-glutamate O4-methyl ester + H(2)O =
 CC protein L-glutamate + methanol.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- DOMAIN: THE N-TERMINAL REGULATORY DOMAIN INHIBITS THE ACTIVITY OF
 CC THE C-TERMINAL EFFECTOR DOMAIN.
 CC -1- PTM: PHOSPHORYLATED BY CHEA. PHOSPHORYLATION SUPPRESSES THE
 CC INHIBITORY ACTIVITY OF THE N-TERMINAL DOMAIN (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 CHEB-TYPE METHYLESTERASE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AB012677; BAA33550.1;
 DR EMBL; AE004575; AAG04848.1;
 DR HSSP; P04042; 1CHD.
 DR InterPro; IPR000673; CheB_methylst.
 DR InterPro; IPR001789; Response_reg.
 DR Pfam; PF00072; response_reg; 1.
 DR Pfam; PF01339; CheB_methylst; 1.
 DR ProDom; PD000039; Response_reg; 1.
 DR ProDom; PD003328; CheB_methylst; 1.
 DR SMART; SM00448; REC; 1.
 DR PROSITE; PS50122; CHEB; 1.
 DR PROSITE; PS50110; RESPONSE REGULATORY; 1.
 KW Hydrolase; Chemotaxis; Sensory transduction; Phosphorylation;
 KW Complete proteome.
 FT DOMAIN 4 121 RESPONSE REGULATORY.
 FT DOMAIN 172 368 CHEB-TYPE METHYLESTERASE.
 FT MOD_RES 55 55 PHOSPHORYLATION (BY SIMILARITY).
 FT ACT_SITE 192 192 BY SIMILARITY.
 FT ACT_SITE 219 219 BY SIMILARITY.

FT ACT_SITE 312 312 BY SIMILARITY.
 FT CONFLICT 34 34 G -> A (IN REF. 1).
 SQ SEQUENCE 368 AA; 39004 MW; FE4801DC220C613B CRC64;
 Alignment Scores:
 Pred. No.: 11.4 Length: 368
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 1 Gaps: 0
 US-10-060-830-1115 (1-60) x CHEB_PSEAE (1-368)
 QY 23 GCTCCAGCGTCCTCGAGCAGC 3
 Db 151 AlaProAlaSerSerSer 157
 RESULT 12
 MDM2_BRARE STANDARD; PRT; 445 AA.
 ID MDM2_BRARE
 AC 042334;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ubiquitin-protein ligase E3 Mdm2 (EC 6.3.2.-) (p53-binding protein
 Mdm2) (Double minute 2 protein).
 GN MDM2.
 OS Brachydanio rerio (zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Neel H., Piette J.;
 RT "Partial cDNA nucleotide sequence of the zebrafish homolog of Mdm2";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BIND P53 PROTEIN AND MAY FUNCTION AS AN UBIQUITIN
 CC LIGASE E3. ZINC IS REQUIRED FOR UBIQUITIN LIGASE E3 ACTIVITY (BY
 CC SIMILARITY).
 CC -1- COFACTOR: ZINC IS REQUIRED FOR UBIQUITIN LIGASE E3 ACTIVITY (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (by similarity).
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -1- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
 CC -1- SIMILARITY: BELONGS TO THE MDM2 / MDM4 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF010255; AAB64176.1;
 DR HSSP; P56273; LYCQ.
 DR InterPro; IPR001515-153; mdm2.
 DR InterPro; IPR003160; MDM2.
 DR Pfam; PF02279; MDM2; 1.
 DR Pfam; PF02279; MDM2; 1.
 DR SMART; SM00184; RING; 1.
 DR PROSITE; PS01358; ZF_RANBP2_1; 1.
 DR PROSITE; PS01358; ZF_RANBP2_2; 1.
 DR PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
 DR PROSITE; PS00518; ZF_RING_2; 1.
 KW Nuclear protein; Ligase; Ubl conjugation pathway; Zinc; Zinc-finger;
 KW Metal-binding.
 FT DOMAIN 12 101 REGION 1.
 FT DOMAIN 141 145 POLY-ARG.
 FT DOMAIN 160 166 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 171 183 NUCLEAR EXPORT SEQUENCE.
 FT DOMAIN 190 279 ARF BINDING.

FT DOMAIN 222 306 REGION II.
 FT DOMAIN 210 276 ASP/GLU-RICH (ACIDIC).
 FT ZN_FING 274 303 RANBP2-TYPE.
 FT ZN_FING 392 433 RING-TYPE.
 FT DOMAIN 420 427 NUCLEOLAR LOCALIZATION SIGNAL
 (POTENTIAL).
 SQ SEQUENCE 445 AA; 49949 MW; 6FA8175A8A8E6261 CRC64;

Alignment Scores:
 Pred. No.: 11.1 Length: 445
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 35.00% Indels: 0
 DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x MDM2_BRARE (1-445)

QY 4 CTGCTCAGAGCTGAGCC 24
 Db 30 LeuLeuGluAspAlaGlyAla 36

RESULT 13
 CPDF_CANFA STANDARD; PRT; 499 AA.
 ID CTGCTCAGAGCTGAGCC 24
 AC Q29473; O02859; PRT; 499 AA.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Cytochrome P450 2D15 (EC 1.14.14.1) (CYP2D15) (P450 DUT2).
 GN CYP2D15
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX TISSUE=Liver;
 RX MEDLINE=95305374; PubMed=7786018;
 RA Sakamoto K., Kirita S., Baba T., Nakanura Y., Yamazoe Y., Kato R.,
 RA Takanaka A., Matsubara T.;
 RT "A new cytochrome P450 form belonging to the CYP2D in dog liver
 RT microsomes: purification, cDNA cloning, and enzyme
 RT characterization.";
 RL Arch. Biochem. Biophys. 319:372-382(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Beagle;
 RX MEDLINE=98162950; PubMed=9504424;
 RA Tasaki T., Nakamura A., Itoh S., Ohashi K., Yamamoto Y., Masuda M.,
 RA Iwata H., Kazusaka A., Kamataki T., Fujita S.;
 RT "Expression and characterization of dog CYP2D15 using baculovirus
 RT expression system.";
 RL J. Biochem. 123:162-168(1998).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98389575; PubMed=9721180;
 RA Roussel F., Duignan D.B., Lawton M.P., Obach R.S., Strick C.A.,
 RA Tweedie D.J.;
 RT "Expression and characterization of canine cytochrome P450 2D15.";
 RL Arch. Biochem. Biophys. 357:27-36(1998).
 CC -1- FUNCTION: HIGH ACTIVITY FOR THE HYDROXYLATION OF BUNITROL AND
 CC IMPIPRINE; LOW ACTIVITY ON DEBRISOQUINE.
 CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) -> ROH +
 CC oxidized flavoprotein + H(2)O.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
 CC -1- TISSUE SPECIFICITY: LIVER. ALSO DETECTED IN SEVERAL OTHER TISSUES.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X07776; CAA30605.1; -.

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D17397; BAA04220.1; -.
 DR HSSP; P00179; LDT6
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; p450; 1.
 DR PRINTS; PR00385; P450.
 KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
 KW Microsome; Endoplasmic reticulum.
 FT INIT_MET 0
 FT BINDING 445 445 HEME (BY SIMILARITY).
 SQ SEQUENCE 499 AA; 56301 MW; 27E352B5B309E7F1 CRC64;

Alignment Scores:
 Pred. No.: 10.9 Length: 499
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x CPDF_CANFA (1-499)

QY 48 GTGTCACATCCATCACCCTTG 28
 Db 310 ValSerThrSerIleThrLeu 316

RESULT 14
 YKP4_KLUULA STANDARD; PRT; 597 AA.
 ID YKP4_KLUULA
 AC P05470;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical killer plasmid pGKL-2 protein 4.
 OS Kluyveromyces lactis (Yeast).
 OG Plasmid pGKL-2.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_TaxID=28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140;
 RX MEDLINE=88289339; PubMed=3041369;
 RA Tommasino S., Ricci S., Galeotti C.L.;
 RT "Genome organization of the killer plasmid pGKL2 from Kluyveromyces
 RT lactis.";
 RL Nucleic Acids Res. 16:5863-5878(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140;
 RA Wilson W.D.;
 RL Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE PRESENCE OF THE TWO LINEAR PLASMIDS, TERMED
 CC PGKL1 AND PGKL2, IN STRAINS OF KLUYVEROMYCES LACTIS CONFERS
 CC THE KILLER PHENOTYPE TO THE HOST CELL, BY PROMOTING THE
 CC SECRETION OF A TOXIN ABLE TO INHIBIT THE GROWTH OF SENSITIVE
 CC STRAINS.
 CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X07776; CAA30605.1; -.

```

DR EMBL; X07946; CAA30772.1; -.
DR PIR; S00962; S00962.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00490; HELIC_C; 1.
KW Hypothetical protein; Plasmid; Hydrolase; Helicase; ATP-binding.
FT NP_BIND 61 68
  ATP (BY SIMILARITY).
FT SITE 146 149
  DEVH BOX.
FT CONFLICT 100 100
  I -> I (IN REF. 2).
FT CONFLICT 575 575
  I -> N (IN REF. 2).
SQ SEQUENCE 597 AA; 69165 MW; 83A6P61583DB8F11 CRC64;

Alignment Scores:
Pred. No.: 10.6 Length: 597
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x YKP4_KLULA (1-597)
QY 30 TTGCTGGCTCCAGGCTCTC 10
Db 358 LeuLeuGlySerValLeu 364
|||||
RESULT 15
HTS_DROME STANDARD; PRT; 1156 AA.
AC Q02645; Q9V8U5; Q9V8U4;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hu-li tai shao protein.
GN HTS OR CG9325
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RX TISSUE=Egg;
RX MEDLINE=94040709; PubMed=1340461;
RA Yue L., Spradling A.C.;
RT 'hu-li tai shao', a gene required for ring canal formation during
FT Drosophila oogenesis, encodes a homolog of adducin.;
RL Genes Dev. 6:2443-2454 (1992).
RN [2]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RX TISSUE=Embryo;
RX MEDLINE=93211992; PubMed=7681599;
RA Ding D., Parkhurst S.M., Lipshitz H.D.;
RT "Different genetic requirements for anterior RNA localization
FT revealed by the distribution of Adducin-like transcripts during
FT Drosophila oogenesis.";
RL proc. Natl. Acad. Sci. U.S.A. 90:2512-2516 (1993).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yeaman M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

```

```

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [4]
RN REVISIONS, AND ALTERNATIVE SPLICING.
RP STRAIN=Berkeley;
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradscky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,
RA Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,
RA Whitfield E.J., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR ASSEMBLING ACTIN AT RING CANALS IN
CC DEVELOPING EGG CHAMBERS. PROBABLY INTERACTS WITH OTHER
CC DEVELOPMENTAL PROTEINS INVOLVED IN NURSE CELL/OOCYTE TRANSPORT
CC THROUGH THE RING CANALS.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND
CC A SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DEVELOPMENTAL STAGE: OOGENESIS AND EARLY EMBRYOGENESIS.
CC -1- MISCELLANEOUS: 'HU-LI TAI SHAO' MEANS 'TOO LITTLE NURSING' IN
CC CHINESE.
CC -1- SIMILARITY: BELONGS TO THE ALDOLASE CLASS II FAMILY. ADDUCIN
CC SUBFAMILY.
CC -1- CAUTION: Ref.3 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L05016; AAA28643.1; -.
CC EMBL; AF151708; AAB59182.1; -.
CC EMBL; AE003796; AAF57565.2; ALT_SEQ.
CC EMBL; AE003796; AAF57566.2; ALT_SEQ.
CC PIR; A46392; A46392.
CC FlyBase; FBgn004873; hts.
CC InterPro; IPR001303; Aldolase_II_N.
CC Pfam; PF00596; Aldolase II; 1.
CC Cytochrome; Membrane; Developmental protein; Alternative splicing.
KW VARSPLIC 659 718
  ALVSLQAKYAFLYSPGQYMYACMKAPLMHKVYVHKVEP
  VSKHNPVPVNDGNMISHN -> GENVONGDSEHSLTFS
  OSSKEFQDVSDGSPKDKKKKGLRTFSLKKKKKRAE
FT FT
FT FT

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 09:28:20 ; Search time 68.9384 Seconds

(without alignments)

11728.275 Million cell updates/sec

Title: US-10-060-830-1113

Perfect score: 3559

Sequence: 1 atgcctctgtctctctgct.....ttttaagaatactcttga 1962

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2.1/USPTO.spool/US10060830/runat_16012003_091239_26852/app_query.fasta_1.6933
-DB=SPREMBL_21 -QFWT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn.1.1.357@runat_16012003_091239_26852 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.rviroid:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3422	96.2	775	4	Q96PD2 homo sapien

2	3422	96.2	775	4	Q8TDX2
3	2984.5	83.9	769	11	Q912V2
4	2950.5	82.9	769	11	Q912V3
5	1213	34.1	364	4	Q14089
6	1151.5	32.4	251	11	Q9D9K5
7	417	11.7	460	11	Q9D696
8	414	11.6	432	11	Q8R327
9	411	11.5	503	11	Q9DAJ3
10	359	10.1	2119	13	Q90X47
11	355.5	10.0	2183	11	Q88783
12	355	10.0	480	4	Q43854
13	352	9.9	480	11	Q35474
14	351	9.9	2224	4	Q43737
15	325	9.1	2343	6	Q18806
16	323.5	9.1	426	11	Q9WFS3
17	321	9.0	463	11	Q9RLX9
18	319	9.0	2343	6	Q62730
19	309.5	8.7	919	13	Q8UVRO
20	309.5	8.7	936	13	Q8UVQ9
21	309	8.7	926	11	Q8QZY7
22	308.5	8.7	216	4	Q14286
23	298	8.4	901	4	Q9H2E4
24	298	8.4	901	4	Q9H2D5
25	298	8.4	906	4	Q9H2E3
26	298	8.4	906	4	Q9H2D4
27	286.5	8.1	923	13	Q8QFX6
28	285.5	8.0	555	4	Q9H2E2
29	283.5	8.0	921	11	Q9QX38
30	271	7.6	363	6	Q77718
31	269.5	7.6	224	11	Q9RLM6
32	267.5	7.5	858	5	Q76470
33	266.5	7.5	609	4	Q96190
34	266.5	7.5	779	4	Q9NTT3
35	265	7.4	644	4	Q961H5
36	265	7.4	704	4	Q9H2E1
37	228	6.4	335	4	Q9BTL9
38	227.5	6.4	1128	11	Q88442
39	224.5	6.3	1158	4	Q14113
40	222.5	6.3	764	11	Q9D2L5
41	222.5	6.3	764	11	Q54860
42	222.5	6.3	764	11	Q8VQD4
43	208	5.8	728	6	Q97567
44	205	5.8	734	4	Q9NUB5
45	205	5.8	734	4	Q96SM3

ALIGNMENTS

RESULT 1

Q96PD2					
ID	Q96PD2	PRELIMINARY:	PRT:	775 AA.	
AC	Q96PD2				
DT	01-DEC-2001	(TRENBLrel. 19, Created)			
DT	01-DEC-2001	(TRENBLrel. 19, Last sequence update)			
DT	01-JUN-2002	(TRENBLrel. 21, Last annotation update)			
DE	Endothelial and smooth muscle cell-derived neuropilin-like protein.				
DE	ESDN.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	PubMed-11447234;				
RA	Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,				
RA	Matsumori A., Sasayama S., Honjo T., Tashiro K.;				
RT	*ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular				
RT	Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is				
RT	Up-regulated after Vascular Injury.;				
RL	J. Biol. Chem. 276:34105-34114(2001).				
CC	-!- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL REGULATORS.				

Db 250 GlnIleSerValIleSerLysGlyThrProTyrTyrGluSerSerLeuAlaAsn 269
QY 451 GTACATCTGTGGGACACTTACTACAAAGTCTTTTACATTAAAGACAAAGTGGATGT 510
Db 270 ValThrSerMetValGlyTyrLeuSerThrSerLeuPheThrPheLysThrSerGlyCys 289
QY 511 TATGGAACACTGGGATGGAGTCTGGTGTGATCGGGATCTCAATAACAGCATCATCT 570
Db 290 TyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSer 309
QY 571 GTGCTGGAGTGGACTGACACACAGAGGCAAGAGAACAGTGGAAACCCAAAAGCCAGG 630
Db 310 ValLeuGluTrpThrAspHisMetGlyGlnGluAsnSerTrpLysProGluLysAlaArg 329
QY 631 CTGAAAAACCTGGACCCGCTGGCTGCTTTTGGCCATGATGATATACCATGGTGTACAA 690
Db 330 LeuArgLysProGlyProProTrpAlaAlaPheAlaThrAspGluHisGlnTrpLeuGln 349
QY 691 ATAGATTTTGAATAAGGAAGAAATAACAGCATATTAACCACTGGATCCACCATGTG 750
Db 350 IleAspLeuAsnLysGluLysIleThrGlyIleValThrThrGlySerThrLeuIle 369
QY 751 GAGCAACAATTACTATGTCTGCTCCACAGAAATCTGTACAGTGTATGGCAGAAATGG 810
Db 370 GluHisAsnTyrTyrValSerAlaTyrArgValLeuTyrSerAspGlyGlnLysTrp 389
QY 811 ACTGTGTACAGAGCCCTGGTGGAGCAAGATAGATATTTCAAGGAAACAAGATATAT 870
Db 390 ThrValTyrArgGluProGlyAlaAlaGlnAspLysIlePheGlnGlyAsnLysAspTyr 409
QY 871 CACCAAGATGCGTAATAACTTTTGGCCCAAAATATTGGCAGCTTTTATTAGAGTGAAT 930
Db 410 HisLysAspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgValAsn 429
QY 931 CTACCCCAATGGCAGCAGAAATTCCTGATGAAATGGAGTCTGCTGGATGTCAGTTATT 990
Db 430 ProValGlnTrpGlnLysIleAlaMetLysValGluLeuLeuGlyCysGlnPheThr 449
QY 991 CTTAAGGTGCTCCCAAACTTACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1050
Db 450 LeuLysGlyArgLeuProLysLeuThrGlnProProProProArgAsnSerAsnAsnLeu 469
QY 1051 AAAACACTACAGCCCTCCAAAATAGCCAAAGTGGTGGCCCAAAATTTACGCAACA 1110
Db 470 LysAsnThrThrValHisProLysLeu-----GlyArgAlaProLysPheThrGlnAla 487
QY 1111 CTACACCTCGCAGTAGCAATGAATTTCTGCACAGACAGAAACAACATGCCAGTCT 1170
Db 488 LeuGlnProArgSerArgAsnAspLeuProLeuProAlaGlnThrThrAlaThrPro 507
QY 1171 GATATCAGAATACTACCTAACTCAATGTAAACCAAGATGTAGCGTGGCTGCAGTT 1230
Db 508 AspValLysAsnThrThrValThrProSerValThrLysAspValAlaLeuAlaVal 527
QY 1231 CTTGTCCCTGCTGCTGATGCTGCTCACTCTCACTCTCACTCTCACTCTCACTCTGCTGG 1290
Db 528 LeuValProValLeuValMetAlaLeuThrThrLeuIleLeuIleLeuValCysAlaThr 547
QY 1291 CACTGGAGAACAGAAAGAAAAAACTGAAGCCACTATGACTTACTTACTGGGACCGG 1350
Db 548 HisTrpArgAsnArgLysLysLysAlaGluGlyThrTyrAspLeuProHisTrpAspArg 567
QY 1351 GCAGTTGTGGAAGAAGATGAGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAA 1410
Db 568 AlaGlyTrpTrpLysGlyValLysGlnLeuLeuProAlaLysSerValGluHisGluGlu 587
QY 1411 ACCCCAGTTCGCTATAGCAGCAGGAGTAAATCACTAGTCCACAGAGAGTCCACCA 1470
Db 588 ThrProValArgTyrSerAsnSerGluValSerHisLeuSerProArgGluValThrThr 607
QY 1471 GTGCTCAGGCTGACTCTGACAGATGATGCTCAGCCACTGGTAGGAGGAATTTGGTACA 1530
Db 608 ValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThr 627

QY 1531 CTTTCATCAAGATCTACCTTTAAACCAAGAAAGAAAGAGCAGGCTATGACAGACTA 1590
Db 628 LeuHisGlnArgSerThrPheLysProGluGlyLysGluAlaSerTyrAlaAspLeu 647
QY 1591 GATCCTTACAACTCACACAGGAGGAGTATATCATCTGCTGACCACTGCAACCACTCCCAATT 1650
Db 648 AspProTyrAsnAlaProValGlnGluValTyrHisAlaTyrAlaGluProLeuProVal 667
QY 1651 ACGGGGCTGAGTATGCAACCCCAATCATGATGACATGTCAGGCGCCACCACTTCA 1710
Db 668 ThrGlyProGluTyrAlaThrProIleValMetAspMetSerGlyHisSerThrAlaSer 687
QY 1711 GTTGGTCAAGCCCTCCACATCCACTTTCAAGGCTAGCGGGAACCAACCTCCCCACTAGTG 1770
Db 688 ValGlyLeuProSerThrSerThrPheArgThrAlaGlyAsnGlnProAlaLeuVal 707
QY 1771 GGAATTTACAACTACACTTCTCCAGGATGACAGCTGCTCCTCAGCGCCAGCCACTAT 1830
Db 708 GlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerGlyGlnAlaGlnTyr 727
QY 1831 GATACCCCGAAGCTGGAGCCAGGCTTCTCCTCCAGGATGACAGTGGTGGTATGCTGATG 1890
Db 728 AspThrProLysGlyGlyLysProAla---AlaAlaProGluGluLeuValTyrGlnVal 746
QY 1891 CCACAGACACACAAGATATCAGGAGCAGGAGGATGGGATGTGATGTTTAA 1950
Db 747 ProGlnSerThrGlnGluAlaSerGlyAlaGlyArgAspGluLysPheAspAlaPheLys 766
QY 1951 GAAATCTCTT 1959
Db 767 GluThrLeu 769
RESULT 4
Q912V3
ID Q912V3 PRELIMINARY: PRT: 769 AA.
AC Q912V3:
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
DE GN ESDN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR.
RX PubMed=11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury";
RL J. Biol. Chem. 276:34105-34114(2001).
CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL: AF387548; AAL30179.1; -.
DR MGD: MGI:1920629; ESDN.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR004043; LCCL_dom.
DR Pfam: PF00431; CUB; 1.
DR Pfam: PF00754; F5_F8_type_C; 1.
DR PROSITE: PS01180; CUB; 1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;
Alignment Scores: 1.41e-244 Length: 769
Pred. No.: 2950.50 Matches: 559
Score:

QY 1891 CCACAGAGCACACAGAACTATCAGACGAGCAGGAAGGATGGGAATGATGTTTAA 1950
 Db 747 ProGlnSerThrGlnGluLeuSerGlyAlaGlyArgAspGluLysPheAspAlaPheLys 766
 QY 1951 GAAATCCTT 1959
 Db 767 GluileLeu 769

RESULT 5
 Q14089 PRELIMINARY; PRT; 364 AA.

AC Q14089;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 40.0 kDa protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shibata T.;
 RL "unpublished";
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL; D29810; BAA18909.1;
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FAS5_C.
 DR InterPro; IPR000403; LCCL_dom.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF00754; F5_F8_type_C; 1.
 DR SMART; SM00042; CUB; 1.
 DR PROSITE; PS01180; CUB; 1.
 KW Hypothetical protein.
 FT NON_TER 1 1
 SQ SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;

Alignment Scores:
 Pred. No.: 1-81e-95 Length: 364
 Score: 1213.00 Matches: 267
 Percent Similarity: 93.47% Conservative: 5
 Best Local Similarity: 91.75% Mismatches: 14
 Query Match: 34.08% Indels: 11
 DB: 4 Gaps: 2

US-10-060-830-1113 (1-1962) x Q14089 (1-364)

QY 67 GGCATACTGTGTCTGGGTTGCAATGAACCAATTCATTAATCAAAAGCAATGAA 126
 Db 76 GlyLysTyrCysGlyLeuGlyLeuGlnIleAsnHisSerIleGluSerLysGlyAsnGlu 95
 QY 127 ATCATTTGCTGTCATGAGTGAATCCATGTTCTGGAGCGGATTTTGGCCTCATAC 186
 Db 96 IleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyr 115
 QY 187 TCTGTTATAGATAACA-AGATCTAATTAATCTGTTTGGACACTGCATCCCAATT-TTTTGG 244
 Db 116 SerValIleAspLysGlnArgSerAsnTyrLeuPheGlyHisCysIleGlnPhePheTrp 135
 QY 245 AACCTGATGTTCACTAGTACTGC-CCAGCTGGTGTCTGCTCTCTCTTTTGGTGAATATCT 303
 Db 136 AsnLeuSerSerValSerThrAlaProAlaGlyCysLeuLeuProPheAlaGluIleSer 155
 QY 304 GGAACAATTCCTCATGATATAGATATCTCGCCATTCGATGCTGCTGTCATGCA 363
 Db 156 GlyThrIleProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAla 175
 QY 364 GGACTAGTGTCAACACAGTTGGGGCGGCAATCAATCACTGTTGTAATTAAGGATTCC 423
 Db 176 GlyValValSerAsnThrLeuGlyGlyGlnIleSerValIleSerLysGlyIlePro 195

QY 424 TATTATGAAGTCTTTGGCTAACACAGTCACATCTGTGGTGGACACTTATCTACAAGT 483
 Db 196 TyrTyrGluSerSerLeuAlaAsnValThrSerValValGlyHisLeuIleLeuGln 215
 QY 484 CTT---TTTACATTTAAACACAAAGTGGATGTTATGAACAACTGGGATGGAGTCTGGTGT- 539
 Db 216 ValPhePheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyGly 235
 QY 540 GATCGGGATCTCAAAATAACAGCATCATCTGTGTGGAGTGGAGTGCACACACAGGGCA 599
 Db 236 AspArgGlySerSerAsnAsnSerIle---ThrValLeuGluTrpThrAspHisThrGlyGln 255
 QY 600 AGAGACAGCTTGAACACCCAAAAGC-CAGGCTGAAAACCTGGACCGCTTGGCGTG 658
 Db 255 nGluAsnSerTrpLysProLysSerGlnAlaGluLysThrTrpThrAlaLeuGlyA 275
 QY 659 CTTTTCGCCACTGATGAATACCAAGTGGTTTACAAATAGATTGAATAGGAAAAAGAAATAA 718
 Db 275 IAPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIle 295
 QY 719 CAGGCATTAATACCACTGGATCCACCATGGTGGAGGACCAATTAATCTATGTCGCTGCTACA 778
 Db 295 hrGlyIleIleThrThrGlySerThrMetVal-SerThrIleThrMetCysLeuProThr 314
 QY 779 GAATCCTGTACAGTGTATGGCAGAAATGGACTGTGTACAGAGAGCCTGGT-GTGGAG 837
 Db 315 GluSerCysThrValMetMetGlyArgAsnGlyLeuCysThrGluSerLeuValValGlu 334
 QY 838 CAGGATAAGATATTTCAAGGAAACAAAGAT---TATCACCAGGATGTCGTATAACTTTT 894
 Db 335 GlnAspLysIlePheGlnGlyAsnLysArgIleIleThrArgMetValArgAsnAsnPhe 354
 QY 895 TTGCCACCAATTAATTCACAGTGGTTTATTTT 921
 Db 355 LeuProIleIleAlaArgLeuLeu 363

RESULT 6
 Q9D9K5 PRELIMINARY; PRT; 251 AA.

AC Q9D9K5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 1700055P21RIK protein.
 GN ESDN OR 1700055P21RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).

```
QY 397 AGTGTGTAATAGTAAGGATTCCTCCTATTATGAAAGTTCTTGGCTAACACGTCACA 456
Db 166 AsnLeuLeuGlnSerLysGlyIleSerHisTyrGluGlyLeuLeuAlaAsnGlyValLeu 185
QY 457 TCTGTGGTGGGACACTATCTACAAGCTTTTACATTTAAAGACAGTGGATGTTATGGA 516
Db 186 SerArgHisGlySerLeuSerGluLysArgPheLeuPheThr 200
QY 517 AACTGGGAGGAGTGTGGTGTGATCGCGGATCCTCAAAATACAGCATCATCTGTGTG 576
Db 200 ----- 200
QY 577 GAGTGGACTGACCACAGGCAAGAGAACAGTGTGAACCCCAAAAGCCAGGCTGAAA 636
Db 200 ----- 200
QY 637 AACCTGGACCGCTTGGGCTGCTTTTGGCACTGATGAATACAGTGGTTACAAATAGAT 696
Db 200 ----- 200
QY 697 TTGAATAAGAAAGAAATAACAGGCATTAATACCACTGGATCCACCATGGTGGAGCAC 756
Db 200 ----- 200
QY 757 AATTACTATGTGCTCGCTACAGAAATCCTCTACAGTGATGATGGGAGAAATGAGCTGTG 816
Db 200 ----- 200
QY 817 TACAGAGAGCGCTGTGTGGAGCAAGATAAGATATTTCAAGGAAACAAAGATTATACCCAG 876
Db 201 -----ProglyMet----- 203
QY 877 GATGTGGTAATAACATTTTGGCCACCAATTAATGTCAGTTTTATTAGAGTGAATCTACC 936
Db 203 ----- 203
QY 937 CAATGGCAGAGAAATTTGCCATGAATAATGGAGCTGCTCGGATGTCAGTTTATTCCTAAA 996
Db 203 ----- 203
QY 997 GGTGCTCTCCTCAAACTTACTCAACCTCCACTCCCTCGGAACAGCAATGACCTCAAAAAC 1056
Db 203 ----- 203
QY 1057 ACTACAGCCCTCCAAAATAGCAAGAGTGTGCCCCAAAATTTACGACCACTACAA 1116
Db 203 ----- 203
QY 1117 CCTGCAGTAGCAATGNAATTCCTGCACAGACAGAACAACTGCCAGTCTGTATATC 1176
Db 203 ----- 203
QY 1177 AGAAATACTACCGTAACCTCAAAATGTAACCAAGATGTAGCGCTGGTGCAGTCTTGTG 1236
Db 204 -----AsnIleThr-----ThrValAlaIleProSerValIle--- 214
QY 1237 CCTGTGCTGTCATGTCCTCACTACTCTCATTTCTCATATTAAGTGTGCTGGCACTGG 1296
Db 215 ---PheIleAlaLeuLeuLeuThrGlyMetGlyIlePheAlaIleCys----- 229
QY 1297 AGAAACAGAAAGAAAGAACTGAAGCCACTATGACTTACTCTTAC-----TGG 1344
Db 230 -----ArgLysArgLysLysGlyAsn-----ProTyrValSerAlaAspAla 244
QY 1345 GACGGGACAGTGTGGTGAAGAAATGAAGCAGTTTCTTCTGCAAAAGCAGTGGACCAT 1404
Db 245 GlnLysThrGlyCysTrpLysGlnIleLysTyr----- 255
QY 1405 GAGGAACCCCA---GTTCCGCTATAGCAGCAGGAA-----GTTATCATCCTGAGTCCA 1455
Db 256 -----ProPheAlaArgHisGlnSerThrGluPheThrIleSerTyrAspAsnGlu 272
```

```
QY 1456 AGAGAAGTCACC-----ACAGTGCTCGAGCTGACTCTGCAGAGTATGTCAGCCA 1506
Db 273 LysGluMetThrGlnLysLeuAspLeuIleThrSerAspMetAlaAspTyrGlnGlnPro 292
QY 1507 CTGTAGGAGGAATCTGTGTACATTCATCAAGATCATCACTTTAAACCA----- 1557
Db 293 LeuMetIleGlyThrGlyThrValAlaArgLysGlySerThrPheArgProMetAspThr 312
QY 1558 -----GAAGAA-----GGAAGAAGACGAGGTATCCACACCTAGATCTTAC 1599
Db 313 AspThrGluGluValArgValAsnThrGluAlaSerGlyHis-----Tyr 327
QY 1600 AACTCACCAGGAGGAAGTTTATCATGCTATGCTGAACCACTCCCAATTTAGGGGCT 1659
Db 328 AspCysProHisArgProGlyArgHisGluTyrAlaLeuProLeuThrHisSerGluPro 347
QY 1660 GAGTATGCAACCCCAATCATC-----ATGGACATG 1689
Db 348 GluTyrAlaThrProIleValGluArgHisLeuLeuArgAlaHisThrPheSerThrGln 367
QY 1690 TCAGGCGACCCCAACACTTCAGTTGTGTCACCTCCACATCCACTTTCAGGCTACGGG 1749
Db 368 SerGlyTyrArgValProGly---ProArgProThrHisLysHisSerHisSerSerGly 386
QY 1750 AACCAACCTCCCCCACTAGTGGGAACCTTACAATACACTTCTCCAGGACTGACAGCTGC 1809
Db 387 GlyPheProAlaThrGlyAlaThrGlnValGlnSerTyrGlnArgProAlaSerPro 406
QY 1810 TCCTCAGCCAGCCCGCAGTATGATACCCGAAAGCTGGGAGCCAGTCTTCTGCCCCA 1869
Db 407 LysProValGlyGlyGlyTyrAspLysPro---AlaAlaSerSerPheLeuAspSerArg 425
QY 1870 GAGCAATTTGTGTACAGGTGCCAGAGACACAGACAGTATCAGGAGCAGGAGGAT 1929
Db 426 Asp-----ProAlaSerGlnSerGlnMetThrSerGlyGlyAspAsp 439
QY 1930 GGG 1932
Db 440 Gly 440
```

RESULT 8

```
Q8R327 PRELIMINARY; PRT; 432 AA.
AC Q8R327;
DT 01-JUN-2002 (TEmBLrel. 21, Created)
DT 01-JUN-2002 (TEmBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEmBLrel. 21, Last annotation update)
DE Similar to RIKEN CDNA 4631413K11 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC026771; AAH2671.1; -
SQ SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD4917 CRC64;
```

Alignment Scores:

Prod. No.:	7-72e-27	Length:	432
Score:	414.00	Matches:	149
Percent Similarity:	32.89%	Conservative:	72
Best Local Similarity:	22.17%	Mismatches:	155
Query Match:	11.63%	Indels:	296
DB:	11	Gaps:	19

US-10-060-830-1113 (1-1962) x Q8R327 (1-432)

```
QY 4 CCTGTGTCCTCGCTCTTACTTGTCTGCTGCTGAGAGCGTGGAGCCAG 63
Db 8 ProSerValLeuAlaLeuLeuPheAlaValCysAlaProLeuArgLeuGlnAlaGluGlu 27
```

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RA MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT *Functional annotation of a full-length mouse cDNA collection.*;
 RL Nature 409:685-690(2001).
 CC -I- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL: AK016485; BAB30265.1; -
 DR MGD: MGI:1913936; 4631413Klikrik.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.
 DR PROSITE: PS01180; CUB; 1.
 SQ SEQUENCE 503 AA; 54547 MW; FEB121E845CAA06B CRC64;

Alignment Scores:

Pred. No.: 1,46e-26 Length: 503
 Score: 411.00 Matches: 147
 Percent Similarity: 32.83% Conservative: 70
 Best Local Similarity: 22.24% Mismatches: 148
 Query Match: 11.55% Indels: 296
 DB: 11 Gaps: 19

US-10-060-830-1113 (1-1962) x Q9DAJ3 (1-503)

QY 37 CTGCTGCTCGAGGAGCTGGAGCCAGCAAGCAATACCTGCTGCTGGGTTCCAAATG 96
 DB 90 LeuLeuPheSerSerAlaThrAspGlnTyrglyProTyrcysgly----SerTrpAlaVal 108
 QY 97 AACCATTCATTAATGAATCAAAAGCAATGAATACATTCATGCTGTCATGAGTGGAAATCCAT 156
 DB 109 ProlysGluLeuArgLeuAsnSerAsnGluValThrValLeuPheLysSerGlySerHis 128
 QY 157 GTTCTGACCGCGGATTTTGGCTCATCTCTCTTATAGTAAACAGATCTAATTAAT 216
 DB 129 IleSerGlyArgGlyPheLeuLeuThrTyraAlaSerSerAspHisProAspLeuIleThr 148
 QY 217 TGTGTGGACACTGCATCAATTTTGGAACTCAGTTCAGTACAGTACGCCAGCTGGT 276
 DB 149 CysLeuGluArgGlySerHisTyrglyPheGluGlySerLysPheCysProAlaGly 168
 QY 277 TGTCTGCTTCCTTTTGTGATATCTGGAACAATTCCTCATGGATATAGATTCCTCG 336
 DB 169 CysArgAspIleAlaGlyAspIleSerGlyAsnThrLysAspGlyTyraArgAspThrSer 188
 QY 337 CCATTTGTCAGTGGTGGTGCATGACAGGAGTAGTGTCAACACAGTGGGGCGGCAATC 396
 DB 189 LeuLeuCysLysAlaAlaIleHisAlaGlyIleIleThrAspGluLeuGlyGlyHisIle 208
 QY 397 ACTGTTGTAATTAAGGTATTCCTTATTAAGAACTTCTTTGGCTTAACAACAGTCA 456
 DB 209 AsnLeuLeuGlnSerLysGlyIleSerHisTyrglyGluGlyLeuLeuAlaAsnGlyValLeu 228
 QY 457 TCTGTGGTGGGACACTTATCTACAAAGTCTTTTACATTTAAGACAAAGTGGATGTTATGGA 516
 DB 336 LeuMetIleGlyThrGlyThrValAlaArgLysGlySerThrPheArgProMetAspThr 355

DB 229 SerArgHisGlySerLeuSerGluLysArgPheLeuPheThr 243
 QY 517 ACACCTGGGATGGAGTCTGGTGTGATCGCGGATCCTCAAAATAACAGCATCATCTGTGCTG 576
 DB 243 ----- 243
 QY 577 GAGTGGACTGACCACACAGGGCAAGAGAACAGTTGGAAACCCCAAAAAAGCCAGCTGAAA 636
 DB 243 ----- 243
 QY 637 AACCTGGACCGCTTGGGCTGCTTTTCCACCTGATGATACCAGTGGTTACAAATAGAT 696
 DB 243 ----- 243
 QY 697 TTGAATAAGAAAAAATAACAGGCATTATAACCACTGGATCCACCATGGTGGAGCAC 756
 DB 243 ----- 243
 QY 757 AATTACTATGTGCTGCTACAGAACTCTACAGTGTATGGGCAGAAATGGAATGACTGTG 816
 DB 243 ----- 243
 QY 817 TACAGAGAGCTGTGTGGAGCAAGATAAGATATTTCAAGGAAACAAGATTAATCACACAG 876
 DB 244 -----ProGlyMet----- 246
 QY 877 GATGTGGTAAATAACTTTTGGCCACCAATTAATTCACAGCTTTTATAGAGTGAATCCTACC 936
 DB 246 ----- 246
 QY 937 CAATGGCAGCAGAAAAATGGCATGAAATGGAGTGTCTGGGATGTCAGTTTATTCCTAAA 996
 DB 246 ----- 246
 QY 997 GGTGCTCTCCAAACCTTACTCAACCTCCACCTCCGGAACAGCAATAGACCTCAAAAAC 1056
 DB 246 ----- 246
 QY 1057 ACTACAGCCCTCCAAAAATAGCCAAAGGTGCTGCCCCCAAAATTTACGCAACCACTACAA 1116
 DB 246 ----- 246
 QY 1117 CCTCGCAGTAGCAATGAATTTCTGTCACAGACAGAAACAACCTGCGAGTCTCTGATC 1176
 DB 246 ----- 246
 QY 1177 AGAAATACTACCGTAACCTCAAAATGTAAACCAAGATGTAGCGCTGGCTGAGTCTTGTCT 1236
 DB 247 -----AsnIleThr---ThrValAlaIleProSerValile--- 257
 QY 1237 CCTGTGCTGCTGCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 1296
 DB 258 ---PheIleAlaLeuLeuLeuThrGlyMetGlyIlePheAlaIleCys--- 272
 QY 1297 AGAAACAGAAAGAAAAAACTGAAGGCACTTACCTGATGATCCTTACCTTACCTTACCTTAC 1344
 DB 273 -----ArgLysArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 287
 QY 1345 GACCGGCAGGTGGTGGAAAGAAATGAAGCAGATTTCTTCTCTGCAAAAGCAGTGGACCAT 1404
 DB 288 GlnLysThrSerCysTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 298
 QY 1405 GAGGAACCCCA---GTTCTGCTATAGCAGCAGCGAA-----GTTATACCTGAGTCCCA 1455
 DB 299 -----ProPheAlaArgHisGlnSerThrGluPheThrIleSerTyraAspAsnGlu 315
 QY 1456 AGAAGATCACCC-----ACAGTGTGCGAGCTGACTGCGAGCTATGCTGAGTCTGCTGAGC 1506
 DB 316 LysGluMetThrGlnLysLeuAspLeuIleThrSerAspMetAlaAspIleGlnGlnPro 335
 QY 1507 CTGTAGGAGGAATTTGTTGGTACACTTTCATCAAAAGATACCTTAAACCA----- 1557
 DB 336 LeuMetIleGlyThrGlyThrValAlaArgLysGlySerThrPheArgProMetAspThr 355

Percent Similarity:	54.31%	Conservative:	28
Best Local Similarity:	40.10%	Mismatches:	75
Query Match:	9.99%	Indels:	15
DB:	11	Gaps:	4
US-10-060-830-1113 (1-1962) x 088783 (1-2183)			
QY	403	GTATTAGTAAGTATTCCTTATGTAAGTCTTTGGCTAACACGTCACATCTGTG	462
DB	1995	IleValAlaArgYrIleAArgIleHisProThrLysSerTyrAsnArgProThr-----	2012
QY	463	GTGGACACTTATCTACAAGCTTTTACATTTTAAGACAAGTGGATGTTATGGAACACTG	522
DB	2013	-----LeuArgLeuGluLeuGlnGlyCysGluValAsnGlyCysSerThrProLeu	2029
QY	523	GGGATGAGTCTGGTGTGATCGGGATCCTCAATATACACATCACTGTGCTGGAG---	579
DB	2030	GlyLeuGluAspGlyArgIleGlnAspLysGlnIleThrAlaSerPheLysSer	2049
QY	580	---TGGACTGACACACAGGCGCAAGAGACAGTGTGGAACCCAAAGACGAGCTGAAA	636
DB	2050	TrpTrpGlyAspYr-----TrpGluProSerLeuAlaArgLeuAsn	2063
QY	637	AAACCTGGA-----CGCCCTGGGCTGCTTTGCCACTGATGAATACCAAGTGTACAA	690
DB	2064	AlaGlnGlyArgValAsnAlaTrpGlnAlaLysAlaAsnAsnLysGlnTrpLeuGln	2083
QY	691	ATAGATTGAATAGGAAAGAAATAACAGGCAATATACCACTGGATCCACCATGGTG	750
DB	2084	ValAspLeuLeuLysIleLysValThrAlaIleValThrGlnGlyCysLysSerLeu	2103
QY	751	GAGCACAATTACTATGTGTCTGCCTACAGAATCTGTACAGTGTATGATGGCGAATGG	810
DB	2104	SerSerGluMetYrValLysSerYrSerIleGlnTyrSerAspGlnGlyValAlaTrp	2123
QY	811	ACTGTGTACAGAGCCTGTGTGGACAGATATTTCAAGGAACAAAGATTTAT	870
DB	2124	LysProYrArgGlnLysSerSerMetValAspLysIlePheGluGlyAsnSerAsnThr	2143
QY	871	CACAGATGCTGCTATACTTTTCCACCAATATTATTCACGCTTTTATAGAGTAAT	930
DB	2144	LysGlyHisMetAsnPheAsnProPheIleSerArgPheIleArgIleIle	2163
QY	931	CCTACCAATGGCAGCAGAAATGGCCATGAAATGAGCTGCTGGATGT	981
DB	2164	ProLysThrTrpAsnGlnSerIleAlaLeuArgLeuGluLeuPheGlyCys	2180
RESULT 12			
ID	O43854	PRELIMINARY; PRT; 480 AA.	
AC	O43854; O43855;		
DT	01-JUN-1998 (TrEMBLrel. 06, Created)		
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)		
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
GN	Integrin-binding protein DEL1 precursor.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.		
RC	TISSUE=EMBRYONIC LUNG;		
RX	MEDLINE=96083109; PubMed=9420328;		
RA	Hidal C., Zupancic T.J., Penta K., Mikhail A., Kawana M.,		
RA	Quertermous E.E., Aoka Y., Fukagawa M., Matsui Y., Platika D.,		
RA	Auerbach R., Hogan B.L.M., Snodgrass R., Quertermous T.,		
RT	"Cloning and characterization of developmental endothelial locus-1: an		
RT	embryonic endothelial cell protein that binds the alphavbeta3 integrin		
RL	receptor.";		
RL	Genes Dev. 12:21-33(1998).		
CC	-1- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH		
CC	INTERACTION WITH THE ALPHA-V BETA3 INTEGRIN RECEPTOR. INHIBITS		

FORMATION OF VASCULAR-LIKE STRUCTURES. MAY BE INVOLVED IN
REGULATION OF VASCULAR MORPHOGENESIS OF REMODELING IN EMBRYONIC
DEVELOPMENT.
-1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
-1- ALTERNATIVE PRODUCTS: TWO ISOFORMS; LONG ISOFORM (SHOWN HERE) AND
SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
-1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
EMBL: U70312; AAC02648.1; -;
DR HSPB; P12259; IC2T.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR000421; FAS6_C.
DR Pfam: PF00008; EGF; 3.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00001; EGF_Like; 2.
DR SMART: SM00231; FAS6C; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; UNKNOWN_2.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS01285; FAS6C_1; 2.
DR PROSITE: PS01286; FAS6C_2; 2.
KW EGF-like domain; Alternative splicing; Signal; Developmental proteol;
Cell adhesion; Repeat; Vascularization.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 480 INTEGRIN-BINDING PROTEIN DEL1.
FT DOMAIN 26 59 EGF-LIKE 1.
FT DOMAIN 78 116 EGF-LIKE 2.
FT DOMAIN 123 134 EGF-LIKE 3.
FT DOMAIN 161 311 F5/8 TYPE C 1.
FT DOMAIN 322 473 F5/8 TYPE C 2.
FT SITE 96 98 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 26 37 BY SIMILARITY.
FT DISULFID 31 48 BY SIMILARITY.
FT DISULFID 50 59 BY SIMILARITY.
FT DISULFID 78 89 BY SIMILARITY.
FT DISULFID 83 105 BY SIMILARITY.
FT DISULFID 107 116 BY SIMILARITY.
FT DISULFID 158 314 BY SIMILARITY.
FT DISULFID 301 305 BY SIMILARITY.
FT DISULFID 319 476 BY SIMILARITY.
FT VARSPPLIC 66 66 A -> G (IN SHORT ISOFORM).
FT VARSPPLIC 67 76 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 480 AA; 53765 MW; F7171E23A309FD48 CRC64;

Alignment Scores:
Pred. No.: 9,26e-22 Length: 480
Score: 355.00 Matches: 75
Percent Similarity: 59.30% Conservative: 27
Best Local Similarity: 43.60% Mismatches: 66
Query Match: 9.97% Indels: 4
DB: Gaps: 2
US-10-060-830-1113 (1-1962) x O43854 (1-480)
QY 472 TTATCTACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACTGGGATGGAG 531
DB 307 LeuArgMetGluLeuLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMeLys 326
QY 532 TCTGGTGTGATCGCGGATCTCTCAATAACAGCATCATCTGTCTGTGGAGTGGACTGACAC 591
DB 327 SerGlyHisIleGlnAspTyrGlnIleThrAlaSerSerIlePheArgThrLeuAsnMet 346
QY 592 ACAGGCAAGAGAACAGTGTGGAACCCAAAGACCCAGCGCTGAAACCACTGGACCG--- 648
DB 347 -----AspMetPheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysVal 364
QY 649 ---CCTTGGGCTCTTTTGGCACTGATACCATCGTGGTGTACAAATAGATTTGAATAAG 705

```

|||||
Db 445 AsnValleAspProPheLeuAlaArgPheIleArgLeuProTrpSerTrpTyr 464
|||
QY 946 CAGAAATCCGATGAAATGGAGCTCTCGGATGT 981
|||||
Db 465 GlyArgIleThrLeuArgSerGluLeuGlyCys 476
|||||
RESULT 14
O43737
ID O43737 PRELIMINARY; PRT; 2224 AA.
AC O43737;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE Factor V.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; Z99572; CAB16748.1; -.
DR HSP; P12259; ICZT.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
SQ SEQUENCE 2224 AA; 251673 MW; 1013108D49690EAB CRC64;

Alignment Scores:
Pred. No.: 3,11e-21 Length: 2224
Score: 351.00 Matches: 77
Percent Similarity: 59.04% Conservative: 21
Best Local Similarity: 46.39% Mismatches: 58
Query Match: 9.86% Indels: 10
DB: 4 Gaps: 3

US-10-060-830-1113 (1-1962) x O43737 (1-2224)

QY 496 AAGCAAGTGGATGTTATGGAACACTGGGATGGATGCTGTGTGATCGCGGATCCAA 555
|||||
Db 2062 GluValAsnGlyCysSerThrProLeuGlyMetGluAsnGlyLysIleGluAsnLysGln 2081
|||||
QY 556 ATAACAGCATCATCTGTCTGGAG-----TGGACTGACCAACACAGGCGCAAGAACAGT 609
|||||
Db 2082 IleThrAlaSerSerPheLysLysSerTrpTrpGlyAspTyr----- 2095
|||||
QY 610 TGGAAACCAAAAGCCAGCTGAAACAACTGGA-----CCGCTTGGCTGCTTTT 663
|||||
Db 2096 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2115
|||||
QY 664 GCCACTGATGAATACCACTGCTTCAATAGATTGTAAGAGAAAAGAAATACACAGGC 723
|||||
Db 2116 AlaAsnAsnLysGlnTrpLeuGluIleAspLeuLeuLysIleLysIleThrAla 2135
|||||
QY 724 ATTATACACTGGATCCACCATGGTGGAGCACAAATTAATGCTGTGCTGCATCAGAAATC 783
|||||
Db 2136 IleIleThrGlnGlyCysLysSerLeuSerGluMetTyrValLysSerTyrThrIle 2155
|||||
QY 784 CTGTACACTGATGATGGCAGCAATGGACTGTGTACAGAGACCTGGTGTGGACAGAT 843
|||||
Db 2156 HistTyrSerGlnGlnGlyValGluTrpLysProTyrArgLeuLysSerSerMetValAsp 2175
|||||
QY 844 AAGATATTTCAAGAAACAAAGATATATACACAGGATGCGGTAATAACTTTTTCACCA 903
|||||
Db 2176 LysIlePheGluGlyAsnThrAsnThrLysGlyHisValLysAsnPheAsnProPro 2195
|||||

```

```

QY 904 ATTATTGCACGTTTATTAGCTGAATCTACCAATGGCAGCAGAAAATTGCCATGAA 963
|||||
Db 2196 IleIleSerArgPheIleArgValIleProLysThrTrpAsnGlnSerIleAlaLeuArg 2215
|||||
QY 964 ATGGAGCTGCTCGGATGT 981
|||||
Db 2216 LeuGluLeuPheGlyCys 2221
|||||
RESULT 15
O18806
ID O18806 PRELIMINARY; PRT; 2343 AA.
AC O18806;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE Factor VIII.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
ON NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Cameron C., Notley C., Hoyle S., McGlynn L., Hough C., Kamisue S.,
RA Giles A., Lallier D.;
RL The canine factor VIII cDNA and 5' flanking sequence.*;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AF016234; AAB87412.1; -.
DR HSP; P00451; ICFG.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265829 MW; A854FAE571C3B99 CRC64;

Alignment Scores:
Pred. No.: 5,39e-19 Length: 2343
Score: 325.00 Matches: 99
Percent Similarity: 43.60% Conservative: 51
Best Local Similarity: 28.78% Mismatches: 121
Query Match: 9.13% Indels: 73
DB: 6 Gaps: 11

US-10-060-830-1113 (1-1962) x O18806 (1-2343)

QY 48 GGAGCTGGAGCCCA---GCAAGGCAATATCTGTGTCTGGGTTGCAAAAT----- 95
|||||
Db 2055 GlyGlnTrpAlaProLysLeuAlaArgLeuHistTyrSerGlySerIleAsnAlaTrpSer 2074
|||||
QY 96 -----GAACATTCAATGAATCAAAAGCAATGAATCAATCATGTCTGTTCATGATGG 149
|||||
Db 2075 ThrLysAspProPheSerTrpIle-LysValAspLeuAlaProMetIleIleHisG1 2094
|||||
QY 150 AATCCATGTTTCGGACCGGATTTTGGCCTCATACTCTCTTTATAGATAACAAGATCT 209
|||||
Db 2094 YlleMetThrGlnGly----- 2099
|||||
QY 210 AATTACTGTTTGGACACATGCATCCAAATTTTGGAAACCTGAGTTCAGTAAGTACTGCC 269
|||||
Db 2100 -----AlaArgGlnLysPheSerSerLeuTyrVa 2109
|||||
QY 270 AGCTGGTGTGCTGCTCTCTCTTTGCTGAGATATCTGGAAACAATCTCTCATGATATAGA 329
|||||
Db 2109 lSerGlnPheIleIleMetTyrSer---LeuAspGlyAsnLysTrpHisSerTyrArgG1 2128
|||||
QY 330 TTCTCTCCCATTTGTCATGCTGTGTGTCATGCAGGAGTAGTGTGCAACACAGCTTGGCGCG 389
|||||

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 09:31:00 ; Search time 13.1166 Seconds
(without alignments)
8802.243 Million cell updates/sec

Title: US-10-060-830-1113

Perfect score: 3559
Sequence: 1 atgcctctgtctctctgct.....tttttaagaacacccctttga 1962

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+_n2p.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10060830/runat_16012003_091240_26877/app_query.fasta_1.6933
-DB=Issued Patents_AA -QFMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS-human40.cdi
-LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn_1_1_50_@runat_16012003_091240_26877 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRAADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	355.5	10.0	2183	3	US-08-746-111-5
2	352	9.9	480	2	US-08-480-229C-10
3	352	9.9	480	2	US-08-659-235C-10
4	352	9.9	513	2	US-08-480-229C-14
5	352	9.9	513	2	US-08-659-235C-14
6	351	9.9	218	1	US-07-607-538C-4
7	351	9.9	218	2	US-08-162-402B-4
8	345	9.7	160	2	US-08-162-402B-14
9	343	9.6	217	1	US-07-607-538C-3
10	343	9.6	217	2	US-08-162-402B-3
11	343	9.6	218	1	US-07-607-538C-2
12	343	9.6	218	2	US-08-162-402B-2

13	343	9.6	387	2	US-08-162-402B-6	Sequence 6, Appli
14	343	9.6	465	2	US-08-162-402B-8	Sequence 8, Appli
15	341	9.6	159	2	US-08-162-402B-12	Sequence 12, Appli
16	330	9.3	321	2	US-08-480-229C-21	Sequence 21, Appli
17	330	9.3	321	2	US-08-659-235C-21	Sequence 21, Appli
18	325	9.1	2343	4	US-09-324-867-2	Sequence 2, Appli
19	323.5	9.1	2304	4	US-09-324-867-4	Sequence 4, Appli
20	322	9.0	2319	1	US-08-212-133A-8	Sequence 8, Appli
21	322	9.0	2319	1	US-08-474-503-6	Sequence 6, Appli
22	322	9.0	2319	2	US-08-670-707A-6	Sequence 6, Appli
23	322	9.0	2319	4	US-09-037-601-6	Sequence 6, Appli
24	322	9.0	2319	4	US-09-315-179-6	Sequence 6, Appli
25	322	9.0	2319	4	US-09-523-656-28	Sequence 28, Appli
26	322	9.0	2319	5	PCT-US94-13200-6	Sequence 6, Appli
27	317	8.9	2351	6	5422260-1	Patent No. 5422260
28	316	8.9	1438	4	US-09-209-916-1	Sequence 1, Appli
29	316	8.9	1471	1	US-08-683-839B-3	Sequence 3, Appli
30	316	8.9	1661	2	US-08-882-083-2	Sequence 2, Appli
31	316	8.9	1661	2	US-08-558-107-2	Sequence 2, Appli
32	316	8.9	1661	4	US-09-243-539-2	Sequence 2, Appli
33	316	8.9	2332	1	US-07-864-004B-4	Sequence 4, Appli
34	316	8.9	2332	1	US-08-251-937A-4	Sequence 4, Appli
35	316	8.9	2332	1	US-08-212-133A-2	Sequence 2, Appli
36	316	8.9	2332	1	US-08-276-594A-2	Sequence 2, Appli
37	316	8.9	2332	1	US-08-474-503-2	Sequence 2, Appli
38	316	8.9	2332	2	US-08-670-707A-2	Sequence 2, Appli
39	316	8.9	2332	4	US-09-037-601-2	Sequence 2, Appli
40	316	8.9	2332	4	US-09-324-867-3	Sequence 3, Appli
41	316	8.9	2332	4	US-09-315-179-2	Sequence 2, Appli
42	316	8.9	2332	4	US-09-523-656-2	Sequence 2, Appli
43	316	8.9	2332	5	PCT-US93-03275-4	Sequence 4, Appli
44	316	8.9	2332	5	PCT-US94-13200-2	Sequence 2, Appli
45	316	8.9	2351	1	US-08-121-202-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-746-111-5
; Sequence 5, Application US/08746111
; Patent No. 6066778
; GENERAL INFORMATION:
; APPLICANT: Ginsburg, David
; APPLICANT: Cui, Jisong
; TITLE OF INVENTION: Compositions And Methods For Screening
; TITLE OF INVENTION: Compounds For Anticoagulant Activity
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,111
; FILING DATE: 06-NOV-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2183 amino acids

```
US-08-659-235C-10
; Sequence 10, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0034-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-659-235C-10

Alignment Scores:
Pred. No.: 8,15e-26 Length: 480
Score: 352.00 Matches: 75
Percent Similarity: 59.30% Conservative: 27
Best Local Similarity: 43.60% Mismatches: 66
Query Match: 9.89% Indels: 4
DB: 2 Gaps: 2

US-10-060-830-1113 (1-1962) x US-08-659-235C-10 (1-480)
QY 472 TTATCTACAAGTCCTTTTACATTTAAGACAGTGGATGATGATGGAACACCTGGGATGGAG 531
Db 307 LeuArgMetGluLeuLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMetLys 326
QY 532 TCTGTGTGTCGCGGATCCTCAATAACAGCATCATCTGCTCGGTGGATGACCTGACAC 591
Db 327 SerGlyHisIleGlnAspTyrGlnIleThrAlaSerValPheArgThrLeuAsnMet 346
QY 592 ACAGGGCAAGACAGTGGAAACCCAAAGCCAGCTGAAACACCTGGACCG-- 648
Db 347 -----AspMetPheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysVal 364
QY 649 ---CCTGTGGTCGCTTTTCCCACTGATGAATACCACTGCTTACAATAGATTTGAATAAG 705
Db 365 AsnAlaTrpThrSerGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuLeuVal 384
QY 706 GAAAGAAATAACAGGCAATTAACCACTGGATCCACCATGCTGGATGGAGACACAAATACAT 765
Db 385 ProThrLysValThrGlyIleIleThrGlnGlyAlaLysAspPheGlyHisValGlnPhe 404

US-08-659-235C-10
; Sequence 10, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0034-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-659-235C-10

Alignment Scores:
Pred. No.: 8,15e-26 Length: 480
Score: 352.00 Matches: 75
Percent Similarity: 59.30% Conservative: 27
Best Local Similarity: 43.60% Mismatches: 66
Query Match: 9.89% Indels: 4
DB: 2 Gaps: 2

US-10-060-830-1113 (1-1962) x US-08-659-235C-10 (1-480)
QY 472 TTATCTACAAGTCCTTTTACATTTAAGACAGTGGATGATGATGGAACACCTGGGATGGAG 531
Db 307 LeuArgMetGluLeuLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMetLys 326
QY 532 TCTGTGTGTCGCGGATCCTCAATAACAGCATCATCTGCTCGGTGGATGACCTGACAC 591
Db 327 SerGlyHisIleGlnAspTyrGlnIleThrAlaSerValPheArgThrLeuAsnMet 346
QY 592 ACAGGGCAAGACAGTGGAAACCCAAAGCCAGCTGAAACACCTGGACCG-- 648
Db 347 -----AspMetPheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysVal 364
QY 649 ---CCTGTGGTCGCTTTTCCCACTGATGAATACCACTGCTTACAATAGATTTGAATAAG 705
Db 365 AsnAlaTrpThrSerGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuLeuVal 384
QY 706 GAAAGAAATAACAGGCAATTAACCACTGGATCCACCATGCTGGATGGAGACACAAATACAT 765
Db 385 ProThrLysValThrGlyIleIleThrGlnGlyAlaLysAspPheGlyHisValGlnPhe 404
```

COUNTRY: USA
ZIP: 94596
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/607,538C
FILING DATE: 01-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Viviana Amzel
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: CRFCC-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 943-1931
TELEFAX: (510) 943-1189
TELEX: N.A.
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE:
US-07-607-538C-4

Alignment Scores:
Pred. No.: 6,89e-26 Length: 218
Score: 351.00 Matches: 77
Percent Similarity: 59.04% Conservative: 21
Best Local Similarity: 46.39% Mismatches: 58
Query Match: 9.86% Indels: 10
DB: 1 Gaps: 3

US-10-060-830-1113 (1-1962) x US-07-607-538C-4 (1-218)

QY 496 AAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGGTGTGATCGCGGATCCTCAA 555
DB 56 GluValAsnGlyCysSerThrProLeuGlyMetGluAsnGlyLysIleGluAsnLysGln 75
QY 556 ATAACAGCATCATCTGCTGCTGGAG-----TGGACTGCACACACAGCGGCAAGAGACACT 609
DB 76 IleThrAlaSerSerPheLysLysSerTrpTrpGlyAspTyr----- 89
QY 610 TGGAAACCCAAAAAGCCAGGCTGAAAAAACCTGGA-----CCGCCTTGGGCTGCTTTT 663
DB 90 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 109
QY 664 GCCACTGTAGTATACAGTGGTTACAAATAGATTGTAATAGAAAGAAATAACAGGC 723
DB 110 AlaAsnAsnLysGlnTrpLeuGluIleAspLeuLysIleLysIleThrAla 129
QY 724 ATTATAACCACTGCATCCACCATGGTGAGCACAATTACTATCTGCTCCCTACAGAAC 783
DB 130 IleIleThrGlnGlyCysLysSerLeuSerSerGluMetTyrValLysSerTyrThrIle 149
QY 784 CTGTACAGTGTATGGGCAAGTGGAGTGTGTGATGAGAGCGCTGGTGTGGAGCAAGAT 843
DB 150 HisTyrSerGluGlnGlyValGluTrpLysProTyrArgLeuLysSerSerMetValAsp 169
QY 844 AGATATTTCAAGGAACAAAGATTATACACAGGATGTCGGTAATTAACTTTTTGGCACCA 903
DB 170 LysIlePheGluGlyAsnThrLysGlyHisValLysAsnPheAsnProPro 189
QY 904 ATTATTGCACGTTTTATTAGTGAATCTTACCATGCGCAGGAGAAATTCGCATGAAA 963
DB 190 IleIleSerArgPheIleArgValIleProLysThrTrpAsnGlnSerIleAlaLeuArg 209
QY 964 ATGAGCTGCTCGGATGT 981

Db 210 LeuGluLeuPheGlyCys 215

RESULT 7

US-08-162-402B-4
Sequence 4, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
TITLE OF INVENTION: GLOBULE (HMEC) ANTIGEN, FRAGMENTS & FUSION PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-4

Alignment Scores:
Pred. No.: 6,89e-26 Length: 218
Score: 351.00 Matches: 77
Percent Similarity: 59.04% Conservative: 21
Best Local Similarity: 46.39% Mismatches: 58
Query Match: 9.86% Indels: 10
DB: 2 Gaps: 3

US-10-060-830-1113 (1-1962) x US-08-162-402B-4 (1-218)

QY 496 AAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGGTGTGATCGCGGATCCTCAA 555
DB 56 GluValAsnGlyCysSerThrProLeuGlyMetGluAsnGlyLysIleGluAsnLysGln 75
QY 556 ATAACAGCATCATCTGCTGCTGGAG-----TGGACTGCACACACAGCGGCAAGAGACACT 609
DB 76 IleThrAlaSerSerPheLysLysSerTrpTrpGlyAspTyr----- 89
QY 610 TGGAAACCCAAAAAGCCAGGCTGAAAAAACCTGGA-----CCGCCTTGGGCTGCTTTT 663
DB 90 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 109
QY 664 GCCACTGTAGTATACAGTGGTTACAAATAGATTGTAATAGAAAGAAATAACAGGC 723

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-2

Alignment Scores:
Pred. No.: 4.16e-25 Length: 218
Score: 343.00 Matches: 71
Percent Similarity: 57.23% Conservativity: 28
Best Local Similarity: 41.04% Mismatches: 68
Query Match: 9.64% Indels: 6
DB: 2 Gaps: 3

US-10-060-830-1113 (1-1962) x US-08-162-402B-2 (1-218)

QY 472 TTATCTACAGTCTTTTACATTTAAGACAGTGGATGTTATGAACACTGGGATGGAG 531
Db 49 LeuArgPheGluLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 68
QY 532 TCTGTGTGTCGGCGATCCTCAATACAGCATCATCTGTCTGGAG---TGGACTGAC 588
Db 69 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerSerTyrlLysThrTrpGlyLeu 88
QY 589 CACACAGGGCAGACAGACAGTGGAAACCCAAAGAGCCAGCTGAAAGAACCTGGA--- 645
Db 89 HisLeu-----PheSerTrpAsnProSerTyrlAlaArgLeuAspLysGlnGlyAsn 105
QY 646 ---CCGCTTGGCTGCTTTTCCACTGATGAATACACAGTGGTTACAAATAGATTTGAAT 702
Db 106 PheAsnAlaTrpValAlaGlySerTyrlGlyAsnAspGlnTrpLeuGlnValAspLeuGly 125
QY 703 AAGGAAAGAAATACAGGCAATATACCACTGATCCACCATGATCCAGTGGGACCAATAC 762
Db 126 SerSerLysGluValThrGlyIleIleThrGlnGlyAlaArgAsnPheGlySerValcIn 145
QY 763 TATGTGTCTGCTCAGACAGTCTGACAGTGTATGATGGCAGAAATGACCTGTCTACAGA 822
Db 146 PheValAlaSerTyrlLysValAlaTrpSerAsnAspSerAlaAsnTrpThrGlyTrpGln 165
QY 823 GAGCCTGTGTGGCAAGATAGATATTTCAAGGAAACAAAGATTTATCACCAGGATGTG 882
Db 166 AspProArgThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisSerHisLys 185
QY 883 CGTAATACTTTTGGCCACCAATTTATGACCTTTTATAGATGCAATCTTACCACATGG 942
Db 186 LysAsnLeuPheGluThrProIleLeuAlaArgTyrlValArgIleLeuProValAlaTrp 205
QY 943 CAGCAGAAATTTGCCATGAAATGCGCTGCTCGGATGT 981
Db 206 HisAsnArgIleAlaLeuArgLeuGluLeuGlyCys 218

RESULT 13

US-08-162-402B-6
Sequence 6, Application US/08162402B
Patent No. 5972337

GENERAL INFORMATION:

APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
GLOBULE (HMF) ANTIGEN, FRAGMENTS & FUSION PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-6

Alignment Scores:
Pred. No.: 5.53e-25 Length: 387
Score: 343.00 Matches: 71
Percent Similarity: 57.23% Conservativity: 28
Best Local Similarity: 41.04% Mismatches: 68
Query Match: 9.64% Indels: 6
DB: 2 Gaps: 3

US-10-060-830-1113 (1-1962) x US-08-162-402B-6 (1-387)

QY 472 TTATCTACAGTCTTTTACATTTAAGACAGTGGATGTTATGAACACTGGGATGGAG 531
Db 218 LeuArgPheGluLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 237
QY 532 TCTGTGTGTCGGCGATCCTCAATACAGCATCATCTGTCTGGAG---TGGACTGAC 588
Db 238 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerTyrlLysThrTrpGlyLeu 257
QY 589 CACACAGGGCAGACAGACAGTGGAAACCCAAAGAGCCAGCTGAAAGAACCTGGA--- 645
Db 258 HisLeu-----PheSerTrpAsnProSerTyrlAlaArgLeuAspLysGlnGlyAsn 274
QY 646 ---CCGCTTGGCTGCTTTTCCACTGATGAATACCACTGGTTTACAAATAGATTTGAAT 702
Db 275 PheAsnAlaTrpValAlaGlySerTyrlGlyAsnAspGlnTrpLeuGlnValAspLeuGly 294
QY 703 AAGGAAAGAAATACAGGCAATTTATCACCAGTGGATCCACCATGTTGGGACCAATAC 762
Db 295 SerSerLysGluValThrGlyIleIleThrGlnGlyAlaArgAsnPheGlySerValcIn 314
QY 763 TATGTGTCTGCTCAGACAGTCTGACAGTGTATGATGGCAGAAATGACCTGTCTACAGA 822
Db 315 PheValAlaSerTyrlLysValAlaTrpSerAsnAspSerAlaAsnTrpThrGlyTrpGln 334
QY 823 GAGCCTGTGTGGCAAGATAGATATTTCAAGGAAACAAAGATTTATCACCAGGATGTG 882
Db 335 AspProArgThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisSerHisLys 354
QY 883 CGTAATACTTTTGGCCACCAATTTATGACCTTTTATAGATGCAATCTTACCACATGG 942
Db 355 LysAsnLeuPheGluThrProIleLeuAlaArgTyrlValArgIleLeuProValAlaTrp 374
QY 943 CAGCAGAAATTTGCCATGAAATGCGCTGCTCGGATGT 981
Db 943 CAGCAGAAATTTGCCATGAAATGCGCTGCTCGGATGT 981

QY 151 ValThrSerValValClyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCys 170
DB 1121 GTCACTTCACGGTGGGATACCTTATCTGCAAGTCTGTTTACATTTAAAGACAAGTGGTTC 1180
QY 171 TyrGlyThrLeuGlyMetGlySerGlyValIleAlaAspProGluInIleThrAlaSerSer 190
DB 1181 TATGGACACTCTGGGATGAGTCTGCTGTGATGTCGATGCCAGATACACATCGTCT 1240
QY 191 ValLeuGluThrThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArg 210
DB 1241 GCATCGAGTGGACTGACCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1300
QY 211 LeuLysLysProGlyProProTrpAlaIlePheAlaThrAspGluTrpGlnTrpLeuGln 230
DB 1301 CTGAGAAACCCGGCCCTCCCTGGGCTGCTTTGGCCACTGATGAGCATCAGTGGCTGCAG 1360
QY 231 IleAspLeuAsnLysGluLysLysIleThrGlyIleThrThrGlySerThrMetVal 250
DB 1361 ATAGACCTTAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1420
QY 251 GluHisAsnTrpThrValSerAlaTrpArgIleLeuThrSerAspGlyGlnLysTrp 270
DB 1421 GAACAGTGTACTATGTCGCTGCTGACAGTCTCTGACAGTGCAGTGGGAGGAGTGG 1480
QY 271 ThrValTrpArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTrp 290
DB 1481 ACTGTGTACAGAACCTGCTGTGGACGAGCAAGATATTTCAAGGAAACAAAGATTAT 1540
QY 291 HisGlnAspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgValAsn 310
DB 1541 CACAAGGATGCTGTAATACTTTTGGCCCACTAATTTGACAGCTTTTCATTTAGAGTGAAC 1600
QY 311 ProThrGlnTrpGlnGlnLysIleAlaMetLysMetLeuLeuGlyCysGlnPheIle 330
DB 1601 CCTGTCAGTGGGCAACAGAAATGGCATGAAAGTGAAGTCTGCGATGTCAGTTTACT 1660
QY 331 ProLysGlyArgProProLysLeuThrGlnProProProProArgAsnSerAsnAspLeu 350
DB 1661 CTCAAAGGTGCGCTTCCAAAGCTTACT-----CCACTCTCGGAACGGCAATAAAGCTC 1714
QY 351 LysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPheThrGlnPro 370
DB 1715 AGAATACTACAGTCTGCTCCCAACTAGTTAAAGTCTGTCGCCCTTAATTTACTCAAGTG 1774
QY 371 LeuGlnProArgSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerPro 390
DB 1775 CTCCAACTCCGAGTAGGATGAATCTCTGTGCGAGCGCGGAGACAACTACCACTCCT 1834
QY 391 AspIleArgAsnThrThrValThrProAsnValThrLysAspValAlaLeuAlaVal 410
DB 1835 GATATAAAAAACACGACTGTAACTCCCACTGTAACCAAGATGTCGCACTGGCTGCCCTT 1894
QY 411 LeuValProValLeuValMetValLeuThrThrLeuIleLeuValCysAlaTrp 430
DB 1895 GTGGTCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1954
QY 431 HisTrpArgAsnArgLysLysLysThrGluGlyThrThrThrAspLeuProTrpTrpAspArg 450
DB 1955 CACTGGAACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2014
QY 451 AlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGlu 470
DB 2015 GCAGGTGTGTTGGAAGGATGAAGCTGCTTCCCTGCGCAAGTCTGCGCAAGGAGGAGG 2074
QY 471 ThrProValArgTrpSerSerSerGluValAsnHisLeuSerProArgGluValThrThr 490
DB 2075 ACGCCAGTCCGTACAGCACTAGTGAAGTCAGTCACTGAGTCCAGGAGGAGGAGGAGGAG 2134
QY 491 ValLeuGlnAlaAspSerAlaGluTrpAlaGlnProLeuValIleGlyIleValGlyThr 510
DB 2135 GTGCTGACGCGGCGACTCTGCAGAAATGACAGCCCTCTGCGGAGGAGTGTGTGCACA 2194
QY 511 LeuHisGlnArgSerThrPheLysProGluGlyLysGluAlaGlyTyrAlaAspLeu 530

DB 2195 CTCCATCAGAGATCCACCTTTAAACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2254
QY 531 AspProTyrAsnSerProGlyGlnGluValTyrHisAlaTrpAlaGluProLeuProIle 550
DB 2255 GATCCTTCAAACTCTCCAACTGACGGAAGTGTACCAAGCTATGCTGAACCACTGCCCTA 2314
QY 551 ThrGlyProGluTrpAlaThrProIleMetAspMetSerGlyHisProThrThrSer 570
DB 2315 ACGGGGCTGTAGTACGCAACCCCGATCTCATGGACATGTCTCAGGCGACCCACACCTCA 2374
QY 571 ValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProProProLeuVal 590
DB 2375 GTTGTCTGCCCTCCACATCCACCTTCAAACTGACGAGGAGGAGGAGGAGGAGGAGGAG 2434
QY 591 GlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTrp 610
DB 2435 GGAACCTTCAACACTCTTCTCTCCAGGACTGACAGCTGTCTCAGGCGACGCTCAGAT 2494
QY 611 AspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTrpGlnVal 630
DB 2495 GACACCCCAAAAGGTGGGAG---TCAGCTGCTACCCCGAGGAGGAGGAGGAGGAGGAG 2551
QY 631 ProGlnSerThrGlnGlnValSerGlyAlaGlyArgAspGlyGluCysAspValPheLys 650
DB 2552 CCCAGAGCACCCAGGAGTATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2611
QY 651 GluIleLeu 653
DB 2612 GAAATCCTT 2620

RESULT 6
AK056350
LOCUS
DEFINITION
to MILK FAT GLOBULE-BGF FACTOR 8 PRECURSOR.
ACCESSION
AK056350.1 GI:16551729
VERSION
oligo capping; fis (full insert sequence).
KEYWORDS
Homo sapiens teratocarcinoma cell line:NT2 CDNA to mRNA,
SOURCE
clone_lib:NT2R12 clone:NT2R12008598.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
AUTHORS
Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
Ono,Y., Hotta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M., Sato,H.,
Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 2020)
Isogai,T., Otsuki,T. and Sugiyama,T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 232-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.

FEATURES
source
1. 2020
/organism="Homo sapiens"

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:52:45 ; Search time 11.6049 Seconds
(without alignments)
1848.294 Million cell updates/sec

Title: US-10-060-830-3
Perfect score: 3888
Sequence: 1 MFLFLLLVLLVLLLEBAGA.....TQVSGAGRGDCDVFKEIL 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	399.5	10.3	901	3	US-08-936-135-22
2	399.5	10.3	906	3	US-08-936-135-24
3	399.5	10.3	909	3	US-08-936-135-8
4	399.5	10.3	909	3	US-08-936-135-10
5	399.5	10.3	914	3	US-08-936-135-12
6	399.5	10.3	926	3	US-08-936-135-14
7	399.5	10.3	931	3	US-08-936-135-16
8	398.5	10.2	925	4	US-09-116-473-2
9	395.5	10.2	909	3	US-08-936-135-18
10	395.5	10.2	926	3	US-08-936-135-20
11	391.5	10.1	923	3	US-08-936-135-6
12	387.5	10.0	922	4	US-09-116-473-4
13	355.5	9.1	2183	3	US-08-746-111-5
14	352	9.1	480	2	US-08-480-229C-10
15	352	9.1	480	2	US-08-659-235C-10
16	352	9.1	513	2	US-08-480-229C-14
17	352	9.1	513	2	US-08-659-235C-14
18	351	9.0	218	1	US-07-607-538C-4
19	351	9.0	218	2	US-08-162-402B-4
20	345	8.9	160	2	US-08-162-402B-14
21	343	8.8	217	1	US-07-607-538C-3
22	343	8.8	217	2	US-08-162-402B-3
23	343	8.8	218	1	US-07-607-538C-2
24	343	8.8	218	2	US-08-162-402B-2
25	343	8.8	387	2	US-08-162-402B-6
26	343	8.8	465	2	US-08-162-402B-8
27	341	8.8	159	2	US-08-162-402B-12

28 330 8.5 321 2 US-08-480-229C-21 Sequence 21, Appl
29 321 8.5 321 2 US-08-659-235C-21 Sequence 21, Appl
30 327.5 8.4 2343 4 US-09-324-867-2 Sequence 2, Appl
31 315 8.1 157 2 US-08-162-402B-13 Sequence 13, Appl
32 315 8.1 463 2 US-08-162-402B-9 Sequence 9, Appl
33 313 8.1 320 2 US-08-480-229C-20 Sequence 20, Appl
34 313 8.1 320 2 US-08-659-235C-20 Sequence 20, Appl
35 311 8.0 1443 2 US-08-670-707A-39 Sequence 39, Appl
36 311 8.0 1443 4 US-09-037-601-39 Sequence 39, Appl
37 311 8.0 1443 4 US-09-315-179-39 Sequence 39, Appl
38 311 8.0 1467 4 US-09-523-656-38 Sequence 38, Appl
39 311 8.0 2133 2 US-08-670-707A-37 Sequence 37, Appl
40 311 8.0 2133 4 US-09-037-601-37 Sequence 37, Appl
41 311 8.0 2133 4 US-09-315-179-37 Sequence 37, Appl
42 311 8.0 2133 4 US-09-523-656-30 Sequence 30, Appl
43 310.5 8.0 2304 4 US-09-324-867-4 Sequence 4, Appl
44 310 8.0 2115 4 US-09-324-867-5 Sequence 5, Appl
45 309.5 8.0 2351 6 5422260-1 Patent No. 5422260

ALIGNMENTS

RESULT 1
US-08-936-135-22
; Sequence 22, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 901 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-936-135-22

Query Match 10.3%; Score 399.5; DB 3; Length 901;
Best Local Similarity 24.1%; Pred. No. 2.3e-26;
Matches 165; Conservative 89; Mismatches 231; Gaps 26;
QY 1 MFLFLLLVLLVLLLEBAGA.....TQVSGAGRGDCDVFKEIL 729
Db 1 MDMFPLTWVFLALYFSGHEVRSQDPPCGGRPNKSDAGYITSPGYPQDPSHQNCWEIVY 60

Qy	59	V-KMGERVRKFG--DDEIDSDSCHFNLYRIYNGIVSRTEIGKYCGGLQMN---HSIE	113
Db	61	APEFNQIVLNFNFHFEIKHD--CKYDFIBIRDGDSADLLGKHG-----NIAPTTII	114
Qy	114	SKGNEITLLEMSGIHVSGRGLASYSVIDKQDLITCLDTASNFLEPFESKYCPAGCLLFP	173
Db	115	SSGSVLVIKTSYAROGAGFSURYEIFTKGS-----EDCSKFTSPNGTIESPG-----F	165
Qy	174	AEISGTHPGYRDSPLCMAGVHAGVYNTLGGQIS-----VVISKGP- 217	
Db	166	PE---KYPHNL-DCTFTILAKPRMEIILQFLTDEHDPLOVCEGDCKYDWLDLDINDGIPH	221
Qy	218	-----YYESSLANNVTSVVGHLSLTSLTFKTS-----GCY 247	
Db	222	VGPLIGYCGTKTPSKLRSSTGLS---LTFHTDMAVAKDGSARYLIHQEPPEFNQC	278
Qy	248	GTLCMESGVTDPOITASSVLEWTDHTGOENSKPKARKLKPGPPWPAFAFDEYQWLQI	307
Db	279	VPLGMESGRITANQISASS-----TFSDGRWTPQOSRLHGDDNGWTPNLDSNKEYLOV	331
Qy	308	DLNKKKITGIITG--STMVEHNYYSARYILYSDGQKWTVYRPPGVEQDKIFQGNKD	365
Db	332	DLAFLTLTAIATOGAISRETQGYVYKYSKYLEVSTNGEDWMVYRHG--KNHKIFQAND	389
Qy	366	YHODVRNRELPPITAEIRVNPQWQOKIAMKMEILLGCQFIPKGRPPKLTQPPPPNSND	425
Db	390	ATEVVLNKLHMPLLTIRFIRPOTWHLGTRALRLLEFG-----RYTDAPCSNMLGN	440
Qy	426	LKNTTAPPKIAGRAPKF-----TOPLOPRSSNEFFPAQTEQTASPDIRNTTVPNTVKD	480
Db	441	LSGLIADTQISASSTREYLWSPSAARLVSSRSGWFFRNPQAQGBEMLQVDLGTPKTVKG	500
Qy	481	V-----ALAAVLVPVLVWVLTLLILLCVAHW-----RNRKKKTEGT--YDLPLW	524
Db	501	VIIQARGGDSITAVEARAFVRKFKYSYSLNGKDWESIQDPRTOQTQKLFEGNMHYDTP--	558
Qy	525	DRAGWNKMKQFLPAKAVDHEETPVYSSSEVNHLSPREVTTVLOADSSEY--AOPLVGG	582
Db	559	-----DIRRFDP-----VPAQYVRVYPERWSPAGIMGRLEVLGCDWTDTSKPTVET	603
Qy	583	IVGLTHQRSTFKP-----EEGKEAG 602	
Db	604	LGPTVKSSEETTPYPMDEDATECG 627	

RESULT 2

```

RESOLUTION 2
US-08-936-135-24
; Sequence 24, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
;

```


QY 174 AEISGTPHGYRDSPLCMAGVHGVSNLTGGQIS-----VVISKGIP- 217
D 166 PE---KYPHNL-DCTFTILAKPRMEIILQFTDLEHDPLOVGECDCKYDWDLDWDGIPH 221
QY 218 -----YYESSLANNVTSVVGHLSTSLTFKTS-----GCY 247
D 222 VGPLIGKYGCTKTPSKLRSSTGILS---LTFHTDMAVAKDGSARYLIHQEPNFQCN 278
QY 248 GTLGMESGVIAADPQTASSVLEWTDHTGOENSWKPKKARLKKPGPPWAAFAFATDEYQWLQI 307
D 279 VPLGMESGRIGNEQISASS-----TFSDGRWTQQSRLHGDGNGWTPNLDNSKEYLQV 331
QY 308 DLNKEKKITGITG--STMVHNHVVSAIRILYSDDGOKWTVYREPGEQDKIFQGNKD 365
D 332 DLRLFTMLTAIATQAGISRETQGYVKSYLEVSTNGEDMVMYRHG--KNHKIFQAND 389
QY 366 YHODVRNNFLPPIIARFIRVNPTQOQKIAMKMLGCGFIPKGRPPKLTQPPPPRNSND 425
D 390 ATEVVLNKLHMLPILTFIRPOTWHLGALRLLEFGC-----RVTDAPCSNMLGM 440
QY 426 LKNTTAPPKIAGRAPK-----TOPQPRSSNEFFPAQTEQTASPDINTTVPNTVKD 480
D 441 LSGLIADTQISASSTREYLWSPSAARLVSSRSGWFPNPAQPGGEWLQVLDLGTPTKTVKG 500
QY 481 V-----ALAAVLVPLVMVLTLLILLYCANHH-----RNRKKKTEGT--YDLPYW 524
D 501 VTIQARGGDSITAVEARAFVRKFKVSYSLNGKDWEYIQDPRTOQTQKLFEGNMHYDTP-- 558
QY 525 DRAGWMKMKOFLPAKAVDHEETPVYSSSEYVHNHLSPREVTVVLOADSAEY--AOPLVGG 582
D 559 -----DIRRDP-----VPAQYVRYVPERWSPAGIMRLEVLGCDWTDTSKPTVET 603
QY 583 IVGTLHQRSTFKP-----EEGKEAG 602
D 604 LGPTVKSEETTPYPMDATECG 627

RESULT 5
US-08-936-135-12
; Sequence 12, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/936,135
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 914 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-936-135-12

Query Match 10.3%; Score 399.5; DB 3; Length 914;
Best Local Similarity 24.1%; Pred. No. 2.4e-26;
Matches 165; Conservative 89; Mismatches 291; Indels 139; Gaps 26;

QY 1 MPLFLLLLVLVLLL--EDAGAOQDGGCHTVLGPESGTLTINYPQTPVPSNVCWEWR 58
D 1 MDMPFLTWVFLALYPSGHEVRSQQPPCGRPNSKDAGYITSPGPQDYPSPHONCEWIVY 60
QY 59 V-KMGERVRIKFG-DFDIEDSDSCHFNILRYINGISVRSIEGKCYCGLQHN--HSIE 113
D 61 APEPNQKVLNPNPHEIEKH-DKYDFIEIRDGSEADLLGKHCG-----NIAPPTII 114
QY 114 SKGNBITLLFMSGIHVSGRGLASVVDKQDLITCLDTASNLFEPFSKYCPAGCLLPF 173
D 115 SSGSVLYIKFTSDYARQAGAGSLRYEIKTGS-----EDCKNFTSPNGTIESPG----F 165
QY 174 AEISGTPHGYRDSPLCMAGVHGVSNLTGGQIS-----VVISKGIP- 217
D 166 PE---KYPHNL-DCTFTILAKPRMEIILQFTDLEHDPLOVGECDCKYDWDLDWDGIPH 221
QY 218 -----YYESSLANNVTSVVGHLSTSLTFKTS-----GCY 247
D 222 VGPLIGKYGCTKTPSKLRSSTGILS---LTFHTDMAVAKDGSARYLIHQEPNFQCN 278
QY 248 GTLGMESGVIAADPQTASSVLEWTDHTGOENSWKPKKARLKKPGPPWAAFAFATDEYQWLQI 307
D 279 VPLGMESGRIGNEQISASS-----TFSDGRWTQQSRLHGDGNGWTPNLDNSKEYLQV 331
QY 308 DLNKEKKITGITG--STMVHNHVVSAIRILYSDDGOKWTVYREPGEQDKIFQGNKD 365
D 332 DLRLFTMLTAIATQAGISRETQGYVKSYLEVSTNGEDMVMYRHG--KNHKIFQAND 389
QY 366 YHODVRNNFLPPIIARFIRVNPTQOQKIAMKMLGCGFIPKGRPPKLTQPPPPRNSND 425
D 390 ATEVVLNKLHMLPILTFIRPOTWHLGALRLLEFGC-----RVTDAPCSNMLGM 440
QY 426 LKNTTAPPKIAGRAPK-----TOPQPRSSNEFFPAQTEQTASPDINTTVPNTVKD 480
D 441 LSGLIADTQISASSTREYLWSPSAARLVSSRSGWFPNPAQPGGEWLQVLDLGTPTKTVKG 500
QY 481 V-----ALAAVLVPLVMVLTLLILLYCANHH-----RNRKKKTEGT--YDLPYW 524
D 501 VTIQARGGDSITAVEARAFVRKFKVSYSLNGKDWEYIQDPRTOQTQKLFEGNMHYDTP-- 558
QY 525 DRAGWMKMKOFLPAKAVDHEETPVYSSSEYVHNHLSPREVTVVLOADSAEY--AOPLVGG 582
D 559 -----DIRRDP-----VPAQYVRYVPERWSPAGIMRLEVLGCDWTDTSKPTVET 603
QY 583 IVGTLHQRSTFKP-----EEGKEAG 602
D 604 LGPTVKSEETTPYPMDATECG 627

RESULT 6
US-08-936-135-14
; Sequence 14, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE

```

; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 926 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-936-135-14

Query Match 10.3%; Score 399.5; DB 3; Length 926;
Best Local Similarity 24.1%; Pred. No. 2.4e-26;
Matches 165; Conservative 89; Mismatches 291; Indels 139; Gaps 26;

Qy 1 MPEFLLLLVLLLL--EDAGAQGGCGHTVLGPESGTLTSINYPTQYPSNVCWEIR 58
Db 1 MDMFPLTWVFLALYFSGHEVRSQDPPCGGRPNKSDAGYITSPGYQDYPHQNCWEIV 60

Qy 59 V-KMGERVRIKFG-DFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYGGLQWN---HSIE 113
Db 61 APEPNQKIVLNFNPFHEKHD-CYDIEIRDGDSEADLLGKHG-----NIAPPTII 114

Qy 114 SKGNEITLLFMSGHVSGRGLASYSVIDKQDLITCLDTASNFLPEFSKYPCAGCLLPF 173
Db 115 SSGSVLYIKFTSDYARQAGFSLRYEIKTGS---EDCSKNFTSPNGTIESPG----F 165

Qy 174 AEISGTHPHGYRDSPLCMAGVHAGVSNLTGGQIS-----VVISKGIP- 217
Db 166 PE---KYPHNL-DCTFTILAKPRMEIILQFLTDLHPDLPQVGECDCKYDWDIDWGI 221

Qy 218 -----YVESLANNVTSVVGHLSTSLFTKTS-----GCY 247

Qy 222 VGPLIGKCGTKTPSKLRSSWGILS---LTHFTDMAVAKDGSARYYLHQEPENFQCN 278
Qy 248 GTLGMSGVIAADPQTASSVLEWTDHTGOENSWKPKARKLKPDPWAAFAFYEQWLQI 307
Db 279 VPLGMSGRIANEQISASS-----TFSDGRWTQQSRLHGDNGWTPNLDNSKEYLQV 331

Qy 308 DLNKKKKTGITTG--STWVEHNYVVSARYLLYSDDGQKWTVYREPGEQDKIFQGNKD 365
Db 332 DLRFUTMLTAITQAGISRETQKGYVKSYLEVSTNGEDMMVYRHG---KNHKIFQAND 389

Qy 366 YHQDVRRNFPPIIARVRNPTQWQKIAMKWEILLGCOFIPKGRPPKLTQPPPPNSND 425
Db 390 ATEVVLNKLHMLPILFRIRPQTHLGLALRELFGC-----RYTDAPCSNMLGM 440

Qy 426 LKNTAPPKIAKRAKPF-----TQLOPRSSNEPPAQTEQTAASPDINDTNTVTPNVTKD 480
Db 441 LSGLIADTQISASSTREYLWSPSAARLVSSRSGWFPNRPQAQPGGEWLQVLDGTPKTVKG 500

Qy 481 V-----ALAAVLVPVLMVLTLLILYCAHHW-----RNRKKTEGT--YDLPYW 524
Db 501 VTIQARGGDSITAVEARAFVRKFKVSYSLNGKDEYIQDPRTOOTKLFEGNMHYDTP--- 558

; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 926 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-936-135-14

Query Match 10.3%; Score 399.5; DB 3; Length 926;
Best Local Similarity 24.1%; Pred. No. 2.4e-26;
Matches 165; Conservative 89; Mismatches 291; Indels 139; Gaps 26;

Qy 1 MPEFLLLLVLLLL--EDAGAQGGCGHTVLGPESGTLTSINYPTQYPSNVCWEIR 58
Db 1 MDMFPLTWVFLALYFSGHEVRSQDPPCGGRPNKSDAGYITSPGYQDYPHQNCWEIV 60

Qy 59 V-KMGERVRIKFG-DFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYGGLQWN---HSIE 113
Db 61 APEPNQKIVLNFNPFHEKHD-CYDIEIRDGDSEADLLGKHG-----NIAPPTII 114

Qy 114 SKGNEITLLFMSGHVSGRGLASYSVIDKQDLITCLDTASNFLPEFSKYPCAGCLLPF 173
Db 115 SSGSVLYIKFTSDYARQAGFSLRYEIKTGS---EDCSKNFTSPNGTIESPG----F 165

Qy 174 AEISGTHPHGYRDSPLCMAGVHAGVSNLTGGQIS-----VVISKGIP- 217
Db 166 PE---KYPHNL-DCTFTILAKPRMEIILQFLTDLHPDLPQVGECDCKYDWDIDWGI 221

Qy 218 -----YVESLANNVTSVVGHLSTSLFTKTS-----GCY 247

Qy 222 VGPLIGKCGTKTPSKLRSSWGILS---LTHFTDMAVAKDGSARYYLHQEPENFQCN 278
Qy 248 GTLGMSGVIAADPQTASSVLEWTDHTGOENSWKPKARKLKPDPWAAFAFYEQWLQI 307
Db 279 VPLGMSGRIANEQISASS-----TFSDGRWTQQSRLHGDNGWTPNLDNSKEYLQV 331

Qy 308 DLNKKKKTGITTG--STWVEHNYVVSARYLLYSDDGQKWTVYREPGEQDKIFQGNKD 365
Db 332 DLRFUTMLTAITQAGISRETQKGYVKSYLEVSTNGEDMMVYRHG---KNHKIFQAND 389

Qy 366 YHQDVRRNFPPIIARVRNPTQWQKIAMKWEILLGCOFIPKGRPPKLTQPPPPNSND 425
Db 390 ATEVVLNKLHMLPILFRIRPQTHLGLALRELFGC-----RYTDAPCSNMLGM 440

Qy 426 LKNTAPPKIAKRAKPF-----TQLOPRSSNEPPAQTEQTAASPDINDTNTVTPNVTKD 480
Db 441 LSGLIADTQISASSTREYLWSPSAARLVSSRSGWFPNRPQAQPGGEWLQVLDGTPKTVKG 500

Qy 481 V-----ALAAVLVPVLMVLTLLILYCAHHW-----RNRKKTEGT--YDLPYW 524
Db 501 VTIQARGGDSITAVEARAFVRKFKVSYSLNGKDEYIQDPRTOOTKLFEGNMHYDTP--- 558

```

Db 222 VGPLIGKCGTKTPSKLRSTGILS---LTFHTDMAVAKDGFSAARYLIHQEPNFCN 278
Qy 248 GTLGMSGVIAPOITASSVLEWTDHTGOENSKPKKARLKKPGPPWAAAFATDEYQWLOI 307
Db 279 VPIGMSGRIGRIANQISASS-----TFSDGRWTPOQSRHGDGNGWTPNLSNKEYLQV 331
Qy 308 DLNKEKKTITG--STMVHNHYVSAYRILYSDGOKWTVYRPGVGEQKIFQGNKD 365
Db 332 DLRFMTLTAITOGAISRETOKGYVVKYKLEVSTNGEDWMVYRHG--KNHKIQAAND 389
Qy 366 YHODVNNFLPPIIARFIRVNPTQMOOKIAMKMLGCOFIPKGRPPKLTQPPPPNSND 425
Db 390 ATEVLNKLHMLPLTFRFIRPQTWHLGIALRLLEFGC-----RVTDAPCSNNLGM 440
Qy 426 LKNTTAPPKIAKGRAPKF-----TOPLQPRSSNEPPAQTEQTASPDIRNTVTPNVTKD 480
Db 441 LSGLIADTQISASTREYLWSPSAARLVSRSGWFPNPOAPGGEWLQVDLGTPTKTVKG 500
Qy 481 V-----ALAAVLVPVLMVLTLLILVCAHW-----RNRKKKTGEGT--YDLPLYW 524
Db 501 VIOGARGGDSITAVEARAFVRKFKVSYSLNGKDWEYIODPRTQOTKLFEGNMHYDTP-- 558
Qy 525 DRAGWKMGMKQFLPAKAVDHEETPVYSSSEVNNHLSPREVTVTLQADSABY--AQPLVGG 582
Db 559 -----DIRREDP-----VPAQYVRVYPERWSPAGIGMRLEVGLGCDWTDKPTVET 603
Qy 583 IVGTLHORSTFKP-----EEGKEAG 602
Db 604 LGPTVKSEETTPYPMDATEDATECG 627

RESULT 8
US-09-116-473-2
; Sequence 2, Application US/09116473
; Patent No. 6428965
; GENERAL INFORMATION:
; APPLICANT: Kolodkin, Alex
; APPLICANT: Ginty, David
; TITLE OF INVENTION: SEMAPHORIN RECEPTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09116,473
; FILING DATE: 17-JUL-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/052,762
; FILING DATE: 17-JUL-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32141
; REFERENCE/DOCKET NUMBER: 01107.74973
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 925 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-116-473-2
Query Match 10.2%; Score 398.5; DB 4; Length 925;
Best Local Similarity 23.7%; Fred. No. 2.9e-26;
Matches 162; Conservative 91; Mismatches 292; Indels 139; Gaps 26;
Qy 1 MFLFLLLLLVLLLL--EDAGAQGDCGCTVLGPESGTLTSINYQTYPNSTVNCWEIR 58
Db 1 MDMPFLTWFLALYSGHKVRSQODPPCCGRLNSKRDAGYITSPGYQDYPHQNCWEVY 60
Qy 59 V-KMGERVRIKFG-DPDIEDSDSCHFNRYRIYNGIVSRTEIGKYCGLGLOMN---HSIE 113
Db 61 AEPNQKIVLNPNPHEIEKHD--CKYDFEIRDGDSESADLLGKHGCG---NIAPPTII 114
Qy 114 SKGNEITLLFMGSIHVSGRGLASVVDKQDLITCLDTASNLEFEEFSKYCPAGCLLP 173
Db 115 SSGSVLYIKTSDYARQGAGFSRLRYEIFTKGS-----EDCSKNFTSPNGTIESPG--F 165
Qy 174 AETSITPHGYRDSPLCMAGVHAGVVSNTLGGQIS-----VVISKGIP- 217
Db 166 PE--KYPHNL--DCFTTILAKPRMEIILQFLTFDLEHDPLOVGECDKYDWLDINDGIPH 221
Qy 218 -----YESSLANNVTSVGHLSLTFKTS-----GCY 247
Db 222 VGPLIGKCGTKTPSKLRSTGILS---LTFHTDMAVAKDGFSAARYLIHQEPNFCN 278
Qy 248 GTLGMSGVIAPOITASSVLEWTDHTGOENSKPKKARLKKPGPPWAAAFATDEYQWLOI 307
Db 279 APLGMSGRIVNEQISASS-----TFSDGRWTPOQSRHGDGNGWTPNVDNKEYLQV 331
Qy 308 DLNKEKKTITG--STMVHNHYVSAYRILYSDGOKWTVYRPGVGEQKIFQGNKD 365
Db 332 DLRFMTLTAITOGAISRETOKGYVVKYKLEVSTNGEDWMVYRHG--KNHKVFAAND 389
Qy 366 YHODVNNFLPPIIARFIRVNPTQMOOKIAMKMLGCOFIPKGRPPKLTQPPPPNSND 425
Db 390 ATEVLNKLHMLPLTFRFIRPQTWHLGIALRLLEFGC-----RVTDAPCSNNLGM 440
Qy 426 LKNTTAPPKIAKGRAPKF-----TOPLQPRSSNEPPAQTEQTASPDIRNTVTPNVTKD 480
Db 441 LSGLIADTQISASTREYLWSPSAARLVSRSGWFPNPOAPGGEWLQVDLGTPTKTVKG 500
Qy 481 V-----ALAAVLVPVLMVLTLLILVCAHW-----RNRKKKTGEGT--YDLPLYW 524
Db 501 VIOGARGGDSITAVEARAFVRKFKVSYSLNGKDWEYIODPRTQOTKLFEGNMHYDTP-- 558
Qy 525 DRAGWKMGMKQFLPAKAVDHEETPVYSSSEVNNHLSPREVTVTLQADSABY--AQPLVGG 582
Db 559 -----DIRREDP-----VPAQYVRVYPERWSPAGIGMRLEVGLGCDWTDKPTVET 603
Qy 583 IVGTLHORSTFKP-----EEGKEAG 602
Db 604 LGPTVKSEETTPYPMDATEDATECG 627

RESULT 9
US-08-936-135-18
; Sequence 18, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 909 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-18

Query Match 10.2%; Score 395.5; DB 3; Length 909;
Best Local Similarity 23.6%; Pred. No. 5.3e-26;
Matches 161; Conservative 92; Mismatches 295; Indels 133; Gaps 25;

QY 1 MFLFLL-LLLVLLLEDAGAGCGGHTVLGPGSGTLTSINYQTPYNSVCEWEIR 58
DB 1 MDMFPLTWFLALYFSRHQVRGQDPDPCGGRNSKADGYITSPGYQDYPHONCEWIVY 60
QY 59 V-KMGERVRKFG-DFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGLGQMN---HSIE 113
DB 61 APEPNQKIVLNFNPHFEIEKHD-CKYDFEIRGDSEADLLGKHG-----NIAPPTII 114
QY 114 SKNETILFLMGSIHVGRGLASYSVIDKDLITCLDTASNFLEPFESKYCPAGCLLPF 173
DB 115 SSGSMLYIKTSDYAROGAGFSLRYEFTKGS-----EDCSKNFTSPNGTIESPG----F 165
QY 174 AEISGTPHGYRSPSSPLCMAGVHAGVVSNTLGGQIS-----VVISKIP- 217
DB 166 PE---KYPHNL-DCTFTILAKPKMEILQFLDLEHDPLOVGECDKYDWDLDWDGIPH 221
QY 218 -----YYESSLANNVTSVGHLSLFT-----FKTSGCYGTL 250
DB 222 VGPLIGKYCTKTPSELRSSTGILSTFTDMAVAKDGFSAARYLVHQPENFCNVPL 281
QY 251 GMSGVIAIDPOITASSVLEWTHGTQENSKPKKARLKPQPPWAAAFATDEYQWLQIDLN 310
DB 282 GMSGRIANEQISASS-----TYSGRWTPQOSRLHGDNDGWTPLNDSNKKEYLOYDLR 334
QY 311 KERKITGIITG--STMVHNHYVSARYILYSDDGQKWTVYRPGVEODKIFQGNKDYHQ 368
DB 335 FLMLTAIATQGAISRETQNGYVVKYKLEVSTNGEDWMYRHG--KNHRKVFQANNDATE 392
QY 369 DVARNELPPIIARFIRVNPQWQOKIAMKMLLGCOFIPKGRPPKLPQPPPRNSDLKN 428
DB 393 VLNKLHAPLITREVRIRPQTHSGIALRLELFC-----RVIDAPCSNMLGSLG 443
QY 429 TTAPPKIAKRAKPF-----TOPLOPRSSNEFFPAQTQOTASPDIRNVTVPNTKDV-- 481
DB 444 LIADSQISASTOETIWMSPSAARLVSSRSGWFFRIPQAQPGEEWLVQDLGTPKTVKGVII 503
QY 482 -----ALAAVLPVLMVLTLLILVCAWHW-----RNRKKKTGST--YDLPYWDRA 527
DB 504 QGARGGDSITAVEARAFVRKFKVSYSLNGKDWEYIDPRTQPKLFEQGNMHYDTP----- 558
QY 528 GWWKGMKQFLPAKAVDHEETPVRYSSSEVNHLSPREVTTLVQADSAEY--AQPLVGGIVG 585
DB 559 ----DIRRFP-----IPAQIVRVYPERWSPAGIGMRLEVLGCDWTDSPKPVKTLGP 606
QY 586 TLHORSTFKP-----EEGKEAG 602

DB 607 TVKSEETTPPYTEEEATECG 627
RESULT 10
US-08-936-135-20
Sequence 20, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
APPLICANT: Chen, Hang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 926 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-20

Query Match 10.2%; Score 395.5; DB 3; Length 926;
Best Local Similarity 23.6%; Pred. No. 5.4e-26;
Matches 161; Conservative 92; Mismatches 295; Indels 133; Gaps 25;

QY 1 MFLFLL-LLLVLLLEDAGAGCGGHTVLGPGSGTLTSINYQTPYNSVCEWEIR 58
DB 1 MDMFPLTWFLALYFSRHQVRGQDPDPCGGRNSKADGYITSPGYQDYPHONCEWIVY 60
QY 59 V-KMGERVRKFG-DFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGLGQMN---HSIE 113
DB 61 APEPNQKIVLNFNPHFEIEKHD-CKYDFEIRGDSEADLLGKHG-----NIAPPTII 114
QY 114 SKNETILFLMGSIHVGRGLASYSVIDKDLITCLDTASNFLEPFESKYCPAGCLLPF 173
DB 115 SSGSMLYIKTSDYAROGAGFSLRYEFTKGS-----EDCSKNFTSPNGTIESPG----F 165
QY 174 AEISGTPHGYRSPSSPLCMAGVHAGVVSNTLGGQIS-----VVISKIP- 217
DB 166 PE---KYPHNL-DCTFTILAKPKMEILQFLDLEHDPLOVGECDKYDWDLDWDGIPH 221
QY 218 -----YYESSLANNVTSVGHLSLFT-----FKTSGCYGTL 250
DB 222 VGPLIGKYCTKTPSELRSSTGILSTFTDMAVAKDGFSAARYLVHQPENFCNVPL 281
QY 251 GMSGVIAIDPOITASSVLEWTHGTQENSKPKKARLKPQPPWAAAFATDEYQWLQIDLN 310
DB 282 GMSGRIANEQISASS-----TYSGRWTPQOSRLHGDNDGWTPLNDSNKKEYLOYDLR 334


```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2183 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-746-111-5

Query Match          9.1%; Score 355.5; DB 3; Length 2183;
Best Local Similarity 40.1%; Pred. No. 7e-22;
Matches 79; Conservative 28; Mismatches 75; Indels 15; Gaps

QY 211 VISKGIPYESSLANNVTSVVGHLSLSTFTKTSCTGLGMSGVIAADPOITASSVLE- 269
Db 1995 IVARYIRIHTPSYKNRPT-----LRLQLGQCEVNCSTPGLSDGRIQDKOITASSPKS 2049

QY 270 -WTDHTGQENSWKPKRKARLKKPG--PPWAAAFATDEYQWLQIDLNKRKKITGIITGSTMV 326
Db 2050 WWDGY-----WEPSLARLNAQGRYNAMQAKANNKQWLQVDLLKIKKVTAIIVTQCKSL 2103

QY 327 EHNYYVSARILYSDGQKWTVYRPGVQEDKIFQGNKDYHQDYVRNNFLPPIIARFTRVN 386
Db 2104 SSEMYKSYSIQYSDQGVAMKPYRQKSMVDKIFEGNSNTRKGMKNFNFPPIISRFIRII 2163

QY 387 PTQWQOKIAMKMLLGC 403
Db 2164 PKTWNQSIARLLEFGC 2180

RESULT 14
US-08-480-229C-10
; Sequence 10, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0026-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 480 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-480-229C-10

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 02:16:00 ; Search time 49.2635 Seconds
(without alignments)
10613.845 Million cell updates/sec

Title: US-10-060-830-1113

Perfect score:

Sequence: 1 atgcctctgttcctcctgct.....ttttaagaaatcctttga 1962

Scoring table: BLOSUM62

pop	10.0	Xgapext	0.5
pop	10.0	Ygapext	0.5
pop	6.0	Fgapext	7.0
pop	6.0	Delext	7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters:	1816940
--	---------

Minimum DB seq length: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Minimum Match	0%
Maximum Match	100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=framed_n2p_model -DEV=xlh
-Q/cgml21/US10060830/runat_16012003_091238_28833/app_query.fasta_1.6933
-DB-A_Genesesd_101002 -QPMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cd1
-LIST=45 -DALIGN=200 -THR_SCORE=pend -THR_MAX=100 -THR_MTN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830_@CG1_1_1230/runat_16012003_091238_28833 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=1 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : A Geneseq 101002:*

1: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1980.DAT: *

2: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1981.DAT: *

3: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1982.DAT: *

4: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1983.DAT: *

5: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1984.DAT: *

6: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1985.DAT: *

7: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1986.DAT: *

8: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1987.DAT: *

9: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1988.DAT: *

10: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1989.DAT: *

11: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1990.DAT: *

12: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1991.DAT: *

13: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1992.DAT: *

14: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1993.DAT: *

15: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1994.DAT: *

16: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1995.DAT: *

17: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1996.DAT: *

18: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1997.DAT: *

19: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1998.DAT: *

20: /SID52/cgcgdata/geneseq/geneseq-emb1/AA1999.DAT: *

21: /SID52/cgcgdata/geneseq/geneseq-emb1/AA2000.DAT: *

22: /SID52/cgcgdata/geneseq/geneseq-emb1/AA2001.DAT: *

23: /SID52/cgcgdata/geneseq/geneseq-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	3369	94.7	669	21	AAV70539	Human Factor 8 Homo	
2	2340	65.7	583	22	AAG75450	Human colon cancer	
3	1375.5	38.6	385	23	AAE22716	Human neuropilin-H	
4	1375.5	38.6	385	23	AAU79460	Human neuropilin-H	
5	1375.5	38.6	385	23	ABB97386	Novel human protei	
6	1322.5	37.2	365	23	AAE22721	Human neuropilin-H	
7	1039	29.2	197	22	ABG01298	Novel human diagno	
8	873.5	24.5	715	22	AAU00670	Human TANGO 229 po	
9	756	21.2	487	22	AAU00628	Novel human protei	
10	756	21.2	539	22	AAU00630	Novel human protei	
11	736	21.2	586	22	AAU00629	Novel human protei	
12	416	11.7	503	21	AAAB19126	Polypeptide isolat	
13	378	10.6	398	23	AAE22715	Human neuropilin-H	
14	378	10.6	398	23	AAU79459	Human Neuropilin-H	
15	355.5	10.0	2183	21	AAAB03533	Murine factor V SE	
16	355	10.0	480	23	AAU98718	Human Integrin bin	
17	352	9.9	480	18	AAW10364	Mouse developmenta	
18	352	9.9	480	20	AAW94683	Murine Del-1 prote	
19	352	9.9	513	20	AAW94685	Murine Del-1 prote	
20	351	9.9	2224	20	AAV49564	Human lipoprotein	
21	349	9.8	159	22	AAAB71907	Human FV C2 domain	
22	347	9.7	2224	17	AAW04254	Human Factor V. H	
23	346	9.7	481	18	AAW10365	Human developmenta	
24	343	9.6	217	16	AAAR7253	HMFG 46 kDa antigen	
25	343	9.6	218	16	AAAR7254	HMFG 46 kDa antigen	
26	343	9.6	387	16	AAAR7252	HMFG 46 kDa antigen	
27	343	9.6	387	21	AAV94453	Human lactadherin	
28	330	9.3	321	20	AAW94697	Human milk fat glo	
29	325	9.1	1431	22	AAE67960	Amino acid sequen	
30	325	9.1	2343	20	AAW80989	Canine factor VIII	
31	325	9.1	2343	21	AAV57846	Canine Factor VIII	
32	323.5	9.1	426	21	AAV94454	Mouse lactadherin	
33	323.5	9.1	2304	21	AAV57848	Mouse Factor VIII	
34	322	9.0	2319	19	AAW53485	Murine factor VIII	
35	322	9.0	2319	19	AAW44135	Mus musculus facto	
36	322	9.0	2319	20	AAV31596	Mouse factor VIII	
37	322	9.0	2319	22	AAE11202	Murine factor VIII	
38	322	9.0	2319	22	AAAB50467	Mouse factor VIII	
39	320	9.0	2332	23	AAU79874	Human factor VIII	
40	317	8.9	2332	23	AAU79875	Human factor VIII	
41	317	8.9	2351	8	AAV70448	Human factor VIII	
42	317	8.9	2351	16	AAV82823	Human factor-VIII	
43	316	8.9	684	16	AAAR73022	Human Factor-VIII	
44	316	8.9	684	16	AAAR74091	Factor-VIII light	
45	316	8.9	1014	8	AAV71139	Factor-VIII:c varia	

ALIGNMENTS

RESULT 1
AAY70539

AA170535
ID AAY70539 standard: Protein: 669 AA:

ID
XX
AC

XX
DT 04-JUL-2000 (first entry)

XX DE Human Factor 8 Homologue.

XX	Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW	cerebroprotective; therapeutic; coagulation related disorder;
KW	haemophilia; stroke; screening.

OS Homo sapiens.

XX
XOMO ЗАРТЕНС.

PN WO200012532-

XX PD

XX 20-AUG-1999; 99WO-US19047.
 XX 31-AUG-1998; 98US-0098521.
 XX (ELIL) LILLY & CO ELI.
 XX Rostock PRJ, Su W, Li XM;
 XX WPI; 2000-256580/22.
 DR N-PSDB; AA251872.
 XX Factor 8 homolog polypeptides and nucleic acids encoding them for
 PT treating coagulation related disorders such as hemophilia and stroke
 XX Claim 3; Page 64-66; 68pp; English.
 XX The present sequence is a human Factor 8 homologue (F8H),
 CC a coagulation cofactor which is selectively expressed in
 CC haematopoietic, heart and reproductive tissues. It has haemostatic and
 CC cerebroprotective activities. The F8H contains a Factor 5/8 signature
 CC and is useful as a therapeutic for treating coagulation related diseases
 CC such as haemophilia and stroke. The nucleic acid is useful as
 CC hybridisation probe and amplification primer for detecting deficiencies
 CC in the level of F8H mRNA, for screening F8H gene mutations and for
 CC monitoring regulation of gene expression. Fragments of the nucleic acid
 CC are also useful as diagnostic probes and primers, and can be used in
 CC screening methods such as those using DNA chips. The present sequence is
 CC also useful as a target to screen therapeutically useful modulators
 CC of the F8H.
 XX SQ Sequence 669 AA;

Alignment Scores:
 Pred. No.: 2,37e-300 Length: 669
 Score: 3369.00 Matches: 631
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 94.66% Indels: 0
 DB: 21 Gaps: 0

US-10-060-830-1113 (1-1962) x AAY70539 (1-669)

QY 67 GCGAATACTGTGCTGGGTTTCAAAATGAACCAATTAATGAATCAAAAGCAATGAA 126
 DB 39 GlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGlu 58
 QY 127 ATCATTTGCTGATGAGTGGGAATCCATGTTTCTGGACCGGATTTTGGCTCATAC 186
 DB 59 IleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyr 78
 QY 187 TCTGTTATAGATAACAGATCTAATTAATTTTGGACACTGCATCCAATTTTTCGAA 246
 DB 79 SerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGlu 98
 QY 247 CTGAGTTTCACTAAGTACTGCCAGCTGGTTGCTGCTTCTTTTGTGATATCTGGA 306
 DB 99 ProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGly 118
 QY 307 ACAATTCCTCATGATATAGATATTCCTGCCATTTGTCATGGCTGGTGCATGCAGGA 366
 DB 119 ThrIleProHisGlyTyrArgAspSerSerProLeucysMetAlaGlyValHisAlaGly 138
 QY 367 GTAGTGTCAACACGTTGGGCGGCAATCATGTTGTAATTAAGTATTTCCTAT 426
 DB 139 ValValSerAsnThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyr 158
 QY 427 TATGAAGTTCTTTGGCTAAACAACCTCATCTGTGTGGGACACTTATCTACAGTCTT 486
 DB 159 TyrGluSerSerLeuAlaAsnValThrSerValValGlyHisLeuSerThrSerLeu 178
 QY 487 TTTACATTAAAGACAGTGGATGTTATGGAACACTGGGGATGGAGTCTGGTGTATCCGC 546

DB 179 PheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAla 198
 QY 547 GATCCTCAAAATAACAGCATCATCTGTCTGGAGTGCACACACAGGCAAGAGAAC 606
 DB 199 AspProGlnIleThrAlaSerSerValLeuGluTyrThrAspHisThrGlyGlnGluAsn 218
 QY 607 AGTTGGAACCCAAAAAGCCAGGCTGAAAAACCTGGACCGCTTGGGCTGCTTTTGGC 666
 DB 219 SerTrpLysProLysLysAlaArgLeuLysLysProGlyProProTrpAlaIlePheAla 238
 QY 667 ACTGATGAATACAGTGGTTACAATAGATTGAATAAGAAAAAGAAATAACAGCAATT 726
 DB 239 ThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGlnLysLysIleThrGlyIle 258
 QY 727 ATAACCACTGGATCCACCATGGTGGACACAAATTACTATGTCTGCTACAGAACTCCTG 786
 DB 259 IleThrThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeu 278
 QY 787 TACAGTGAATGGGAGAAATGGACTGTGTACAGAGAGCCTGGTGGAGCAAGATAAG 846
 DB 279 TyrSerAspAspGlyGlnLysTrpThrValTyrArgGluProGlyValGluGlnAspLys 298
 QY 847 ATATTTCAAGGAACAAAGATTATCACAGAGTGTGGTAATAACTTTTGGCCACCAATT 906
 DB 299 IlePheGlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProIle 318
 QY 907 ATTGCACGTTTATTAGAGTGAATCTCCCAATGGCAGCAGAAAAATTCGCATGAAATG 966
 DB 319 IleAlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMet 338
 QY 967 GAGCTCTCGGATGTCAGTTTATTCCTAAAGTGTCTCTCAAAACTTACTCAACCTCCA 1026
 DB 339 GluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProPro 358
 QY 1027 CTTCTCTCGGACAGCAATGACCTCAAAACACACTACACCCCTCCCAAAATAGCCAAAGT 1086
 DB 359 ProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGly 378
 QY 1087 CGTGCCCAAAATTTACGCAACCACTACAACCTCGCAGTAGCAATGAATTTCTGTCACAG 1146
 DB 379 ArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGln 398
 QY 1147 ACAGAACAAACACTGCCAGTCTGTATCAGAAATACTACCGTAACCTCAAAATGTAACT 1206
 DB 399 ThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThr 418
 QY 1207 AAAGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1266
 DB 419 LysAspValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThrThrLeu 438
 QY 1267 ATTCTCATATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1326
 DB 439 IleLeuIleLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysLysLysLysLys 458
 QY 1327 TATGACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 1386
 DB 459 TyrAspLeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuPro 478
 QY 1387 GCAAAACAGTGGACCATGAGAAACCCACCTGCTGCTATAGCAGCAGGCAAGTAAATCAC 1446
 DB 479 AlaLysAlaValAspHisGluGluThrProValArgTyrSerSerSerGluValAsnHis 498
 QY 1447 CTGAGTCCAAAGAGAAAGTCAACACAGTGTGCTGAGGCTGACTGTCAGAGTATGCTCAGCA 1506
 DB 499 LeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyrAlaGlnPro 518
 QY 1507 CTGCTAGAGGAATTTGGTGTACTTCAATCAAGATCTACTTTAAACCCAGAGAAAGCA 1566
 DB 519 LeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGly 538
 QY 1567 AAAGACAGCGCTATGACAGACCTAGATCCTTACAACTCACAGGCGCAGGAGATTTATCAT 1626
 DB 539 LysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHis 558

Db 285 GlnIleThrAlaSerValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrp 304
 QY 613 AACCCAAAAGCCAGCGTGAACACCTGGACCGCTGGCTGTTTGGCCACTGAT 672
 Db 305 LysProLysLysAlaArgLeuLysLysProGlyProProTrpAlaAlaPheAlaThrAsp 324
 QY 673 GAATACAGTGGTTACAAATAGATTTGAATAAGGAAGAAATAACAGCATTAATAC 732
 Db 325 GluTyrGlnIlePheGlnIleAspLeuAsnLysGluLysLysLeuThrGlyIleIleThr 344
 QY 733 ACTGGATCCACCATGGTGGAGCAATTAATGCTGCTCCCTACAGAAATCCCTGTACAGT 792
 Db 345 ThrGlyIleThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSer 364
 QY 793 GATGATGGGAGAAATGGATGCTGTACAGAGACGCTGTGTGGAGCAAGATAGATATT 852
 Db 365 AspAspGlyGlnLysTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePhe 384
 QY 853 CAAGGAACAAAGATATACACAGATGTCGTAATAACTTTTGGCCACCAATTAATGCA 912
 Db 385 GlnGlyAsnLysAspTyrHisGlnAspValArgAsn**PheLeuProProlIleIleAla 404
 QY 913 CGTTTTATTAGAGTGAATCTACCAATGGCAGCAGAAATGCCATGAAATGGAGCTG 972
 Db 405 ArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeu 424
 QY 973 CTCGGATGTCAGTTTATCTCTAAAGTGGCTCCCAAACTTACTCAACCTCCACCTCT 1032
 Db 425 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProPro 444
 QY 1033 CGGAACAGCAATGACCTCAAAACACATACAGCCCTCCAAATAGCAAGGTCGTGCC 1092
 Db 445 ArgAsnSerAsnAspLeuLysAsnThrAlaProProlLysIleAlaLysGlyArgAla 464
 QY 1093 CCAAAATTTAGCCACCACTACACCTCGCAGTAGCAATGAATTCCTGCACAGACAGAA 1152
 Db 465 ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 484
 QY 1153 CAACACTGCGCTGATCAGAAATACCTACCTCACTCAATGTCACCAATGTAACCAAGAT 1212
 Db 485 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 504
 QY 1213 GTAGCGTGGCTGAGTTCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1272
 Db 505 ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 524
 QY 1273 ATATTAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1332
 Db 525 IleLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrAsp 544
 QY 1333 TTACTCTTACTGGACCGGCGAGGTGGTGGAAAGGAATG 1371
 Db 545 LeuProTyrTrpAspArgAlaGlyAsnSerArgGlyLeu 557
 RESULT 3
 AAEE22716
 ID AAEE22716 standard; Protein: 385 AA.
 AC
 XX AAEE22716;
 XX
 DT 09-AUG-2002 (first entry)
 XX
 DE Human neuropilin-Hy2 protein.
 XX
 KW Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
 KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 KW thrombocytopoiesis; memory; platelet; plastic anaemia; antiinflammatory;
 KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;

KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 KW nootropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
 KW cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
 KW immunosuppressive; chromosome 6q21.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..20
 FT /label= Signal_peptide
 FT Protein 61..385
 FT /note= "Human mature neuropilin-Hy2 protein"
 XX
 XX WO200222815-A1.
 PD 21-MAR-2002.
 XX
 PF 12-SEP-2001; 2001WO-US28488.
 XX
 PR 11-SEP-2000; 2000US-0659671.
 PR 06-SEP-2001; 2001US-317902P.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT;
 XX
 XX WPI; 2002-393966/42.
 DR N-PSDB; AAD35994.
 DR
 XX
 PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 PT useful for treating neurodegenerative diseases e.g. Alzheimer's
 PT disease, and for diagnosing and mapping genetic neuronal defects
 XX
 PS Claim 3; Page 128-130; 152pp; English.
 CC
 CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 CC like polypeptides and polynucleotides are useful in modulating neuronal
 CC growth regenerative capacity, treating neurodegenerative diseases,
 CC diagnosing and mapping genetic neuronal defects and degenerative diseases.
 CC like Alzheimer's disease and for treating learning and memory disorders.
 CC They are also useful for inducing angiogenesis, neovascularisation, as
 CC well as organ growth and development e.g. heart and other tissues.
 CC Antagonists of neuropilin-like polypeptides are useful for treating
 CC cancers and other malignant diseases. Neuropilin is used to treat
 CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 CC nocturnal haemoglobinuria and is used in nerve tissue growth or
 CC regeneration, in wound healing, tissue repair and replacement and in
 CC healing of bones, incisions and ulcers. Compositions comprising the
 CC sequences of the invention are useful for treating diseases of peripheral
 CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 CC stroke, ulcers, immune deficiencies and immune disorders, infections by
 CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 CC inflammatory eye disease. The nucleic acids of the invention are used in
 CC gene therapy techniques. The present sequence is human neuropilin-Hy2
 CC protein. Neuropilin-Hy2 gene is located on chromosome 6q21.
 XX
 SQ Sequence 385 AA;
 Alignment Scores:
 Pred. No.: 2.07e-117 Length: 385
 Score: 1375.50 Matches: 296
 Percent Similarity: 77.26% Conservative: 3
 Best Local Similarity: 76.49% Mismatches: 8
 Query Match: 38.65% Indels:

PI Tang YT;
XX WPI; 2002-3939566/42.
XX Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects
XX
PS Disclosure; Page 131-132; 152pp; English.
XX The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.
CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat
CC platelet disorders e.g. thrombocytopenia, plastic anaemia and paroxysmal
CC nocturnal haemoglobinuria and is used in nerve tissue growth or
CC regeneration, in wound healing, tissue repair and replacement and in
CC healing of bones, incisions and ulcers. Compositions comprising the
CC sequences of the invention are useful for treating diseases of peripheral
CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-brager
CC syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
CC stroke, ulcers, immune deficiencies and immune disorders, infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy2
XX mature protein.
XX
XX Sequence 365 AA;
Alignment Scores:
Pred. No.: 1.49e-112 Length: 365
Score: 1322.50 Matches: 274
Percent Similarity: 95.85% Conservative: 3
Best Local Similarity: 94.81% Mismatches: 8
Query Match: 37.16% Indels: 7
DB: 23 Gaps: 1
US-10-060-830-1113 (1-1962) x AAE22721 (1-365)
Qy 67 GCGAATCTGTGCTGGGTTGCAATGAACCATTCATTAATGAATCAAAAGGCAATGAA 126
Db 79 GlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGlu 98
Qy 127 ATCACTGCTGTCATGAGTGGATCCATCTTTCTGACGCGGATTTTGGCTCATAC 186
Db 99 IleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyr 118
Qy 187 TCTGTTATAGATAACAAGATCTAATTAATCTTTTGGACACTGCATCCAAATTTTGGAA 246
Db 119 SerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGlu 138
Qy 247 CCTGAGTTCAGTAACTAGTATAGATTCCTCGCATTTGTCATGGTGGTGCATGATCTGGA 306
Db 139 ProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGly 158
Qy 307 ACAATTCCTCATGATATAGATTCCTCGCATTTGTCATGGTGGTGCATGCGAGA 366
Db 159 ThrIleProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGly 178
Qy 367 GTAGTGTCAACACAGCTTGGGGCCAAATCAGTGTGTAATTAAGTAAGTATTCCTCAT 426
Db 179 ValValSerAsnThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyr 198

QY 427 TATGAAGATTCTTGGCTAACACGTCACATCTGTGTGGGACACTTATCTACAAGTCTT 486
Db 199 TyrGluSerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeu 218
QY 487 TTTACATTTTAAGACAAGTGGATCTTATGGAACACTGGGGATGGAGTCTGTGTGATCGC 545
Db 219 PheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyGlyAspArg 238
QY 546 GGATCTCAAAATAACAGCATCATCTGTGTGGAGTGGACTGACACACAGGGCAAGAGAA 605
Db 239 GlySerSerAsnAsnSerIle-ThrValLeuGluTrpThrAspHisThrGlyGlnGluAs 258
QY 606 CAGTGTGAACACCAAAAGC-CAGGCTGAAAAACCTGGACCCCTTGGCTCTTTTG 664
Db 258 nSerTrpLysProLysSerGlnAlaGluLysThrTrpThrAlaLeuGly-AlaPheA 278
QY 665 CCACTGATGAATACCACTGGTGTACAAATAGATTGAATAAGGAAAAAGAAATAACAGCA 724
Db 278 lathraspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysIleThrGlyI 298
QY 725 TTAATCACTGGATCCACCATGGTGGAGCACAAATTACTATGTCTCTGCCTACAGATCC 784
Db 298 leileThrThrGlySerThrMetVal-SerThrIleThrMetCysLeuProThrGluSer 317
QY 785 TGTACAGTGTGATGGGCAGAAATGGACTGTGTACAGAGAGCCCTGGT-GTGGAGCAAGAT 843
Db 318 CysThrValMetMetGlyArgAsnGlyLeucysThrGluSerLeuValValGluGlnAsp 337
QY 844 AAGATATTTCAAGGAAACAAAGAT---TATCACCAAGGATGTGCGTAATAACTTTTGGCA 900
Db 338 LysIlePheGlnGlyAsnLysArgIleIleThrArgMetValArgAsnAsnPheLeuPro 357
QY 901 CCAATTTATTCACGTTTATT 921
Db 358 ProIleIleAlaArgLeuLeu 364
RESULT 7
ABG01298
ID ABG01298 standard; Protein; 197 AA.
XX AC ABG01298;
XX DT 13-FEB-2002 (first entry)
XX DX
XX DE Novel human diagnostic protein #1289.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS65485.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity

PS Claim 20; SEQ ID No 31657; 103pp; English.

XX. The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 197 AA;

Alignment Scores:
Pred. No.: 1,25e-86 Length: 197
Score: 1039.00 Matches: 197
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.19% Indels: 0
DB: 22 Gaps: 0

US-10-060-830-1113 (1-1962) x ABG01298 (1-197)

QY 1369 ATGAAGCAGTTCCTTCGCAAAAGCAGTGGACCATGAGGAACCCAGTTCGCTATAGC 1428
Db 1 MetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSer 20
QY 1429 AGCAGCGAAGTTATCACTCAGTCCAAAGAGAGTCCACAGTGTGCAGGCTCAGCTCT 1488
Db 21 SerSerGluValAlaSerHisLeuSerProArgGluValThrValLeuGlnAlaAspSer 40
QY 1489 GCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTGTACACTTCATCAAGATCTACC 1548
Db 41 AlaGluTyrAlaGlnProLeuValGlyGlyLeuValGlyThrLeuHisGlnArgSerThr 50
QY 1549 TTTAAACAGAGAGAGGAAAGAGAGCAGCTATGACAGCTAGATCCTTACAACTACCA 1608
Db 61 PheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerPro 80
QY 1609 GGGCAGGAGTTCATGCTATGCTGCTCAACCACTCCCAATACGGGCTGAGTATGCA 1668
Db 81 GlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAla 100
QY 1669 ACCCAATCATCATGACATGTGAGGACCCCACTTCAGTTGGTCCAGCCCTCCACA 1728
Db 101 ThrProIleLeuMetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThr 120
QY 1729 TCCACTTTCAAGGCTACGGGAACCACTCCGCCACTAGTGGGAACCTACATACACTT 1788
Db 121 SerThrPheLysAlaThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeu 140
QY 1789 CTCCTCAGAGTACAGCTGCTCTCAGCCAGGCCCCAGTATGATACCCGGAAGCTGGG 1848
Db 141 LeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGly 160
QY 1849 AAGCCAGCTCTACCTGCCCGCAGACCAATGGTGTACCAAGTCCAGAGTCCACAGACACA 1908
Db 161 LysProGlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGlu 180
QY 1909 GTATCAGGAGCAGGAGGATGGGGAATGTGATGTTTAAAGAAATCCTT 1959

Db 181 ValSerGlyAlaGlyArgAspGlyGluCysAspValPheLysGluLeu 197
RESULT 8
AAU00670
ID AAU00670 standard; Protein: 715 AA.
XX AC AAU00670;
XX DT 07-SEP-2001 (first entry)
XX DE Human TANGO 229 polypeptide.
XX KW Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung; skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte; bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer; anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis; attention deficit disorder; Crohn's disease; gastroenteritis; goitre; hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism; muscular dystrophy; immuno-competence; vertebrate; blood; serum.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Peptide 1..34
XX FT Domain /note= "Signal peptide"
XX FT Protein /note= "Extracellular domain"
XX FT Domain /note= "Mature human TANGO 229"
XX FT Domain /note= "Transmembrane domain"
XX FT Domain /note= "Cytoplasmic domain"
XX PN WO200129088-A1.
XX PD 26-APR-2001.
XX PF 23-JUN-2000; 2000WO-US17386.
XX PR 19-OCT-1999; 99US-0420707.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PI Mackay CR, Myers PS, KIRST SJ, Fraser CC, Leiby KR;
XX WP1; 2001-308477/32.
XX N-PSDB; AAS00660.
XX PT New isolated nucleic acid molecule for diagnosis, prevention, and therapy of human and other animal disorder, or as modulating agent for regulating cellular processes -
XX PS Claim 8; Fig 1; 263pp; English.
XX CC The sequence represents human TANGO 229 polypeptide. This protein and similar others exhibit the ability to affect growth, proliferation, survival, differentiation, activity, morphology, or movement/migration of, e.g. T cells and cells of the heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus tissue. They can be used as modulating agents for regulating cellular processes, thus, the proteins and their associated nucleic acids can be used to prognosticate, prevent, diagnose, or treat disorders associated with physiological processes. These disorders include abnormal blood coagulation, asthma, anaphylaxis, hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention deficit disorder, Crohn's disease, gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and muscular dystrophy. Antibodies to disorders such as these can be made by providing a polypeptide of the invention to an immuno-competent

CC vertebrate and harvesting blood or serum from the vertebrate.

XX Sequence 715 AA;

Alignment Scores:

Pred. No.: 3,26e-71 Length: 715
Score: 873.50 Matches: 225
Percent Similarity: 51.17% Conservative: 102
Best Local Similarity: 35.21% Mismatches: 209
Query Match: 24.54% Indels: 103
DB: 22 Gaps: 22

US-10-060-830-1113 (1-1962) x AAU00670 (1-715)

QY 37 CTGCTGCTGAGACGCTGGAGCCAGCAGCAATCTACTGTGGTCTGGGGTGCAGATG 96
DB 99 LeuLeuPheThrSerSerSerAspGlnTyrglyProTyrcysgly---SerMetThrVal 117
QY 97 AACCAATTCATTAAGTCAAAAGGCAATGAATACATTCATCTGCTCATGAGTGAATCCAT 156
DB 118 ProLysGluLeuLeuLeuAsnThrSerGluValThrValArgPheGluSerGlySerHis 137
QY 157 GTTCTGCGAGCGGATTTTGGCCCTATCTCTGTATAGATAAACAAGATCTAATTACT 216
DB 138 IleSerGlyArgGlyLeuLeuLeuThrTyraLeuSerSerAspHisProAspLeuIleThr 157
QY 217 TGTTCGACACTGCATCCAAATTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGGT 276
DB 158 CysLeuGluArgAlaSerHisTyrlsLeuLysThrGluTyrlsLeuPheCysProAlaGly 177
QY 277 TGCTCTGCTTCTCTGCTGATATCTGGAACAATTCCTCATGGATATAGATTCCTCG 336
DB 178 CysArgAspValAlaGlyAspIleSerGlyAsnMetValAspGlyTyraArgSer 197
QY 337 CCATTGTGATGCTGGTGTGATGAGGAGTGTCTCAACACAGTGTGGCGGCCCAATC 396
DB 198 LeuLeuCysLysAlaAlaIleHisAlaGlyIleAlaAspGluLeuGlyGlyGlnIle 217
QY 397 AGTGTGTGATTAAGTATTCCTCTATTATGAAGTCTTGGCTAACACGTCACA 456
DB 218 SerValLeuGlnArgLysGlyIleSerArgTyrgluGlyLeuLeuAlaAsnGlyValLeu 237
QY 457 TCTGTGTGGACACTTATCTACAGTCTTTTACATTTTAAGACAAGTGGATTTATGGA 516
DB 238 SerArgAspGlySerLeuSerAspLysArgPheLeuPheThrSerAsnGlyCysSerArg 257
QY 517 ACATGGGATGAGGCTGTGTGTGATCGCGGATCTCAATAACACATCATCTGTGCTG 576
DB 258 SerLeuSerPheGlu-----ProAspGlyGlnIleArgAlaSerSerTrp 273
QY 577 GAGTGGACTCACACACAGGCGAAGACAGTGTGAACCCAAACCCAGCGCTGAAA 636
DB 274 GlnSerValAsnGluSerGlyAspGlnValHisTrpSerProGlyGlnAlaArgLeuGln 293
QY 637 AAACCTGGACCGCTTGGGCTGTGTGCTGCTACAGATCTCTACAGTGTGATGGCAG 684
DB 294 AspGlnGlyProSerTrpAlaSerGlyAspSerSerAsnHisLysProArgGluTrp 313
QY 685 TTCAATAATGATTGATAGGAAGAAGAAATAACAGGCGATTAACACTGATCCAC 744
DB 314 LeuGluIleAspLeuGlyGluLysLysLysIleThrGlyIleArgThrThrGlySerThr 333
QY 745 ATGGTGGACACAAATTAATCTGTCTGCTGCTACAGATCTCTACAGTGTGATGGCAG 804
DB 334 GlnSerAsnPheAsnPheTyrlsValLysSerPheValMetAsnPheLysAsnAsnSer 333
QY 805 AAATGAGTGTGTACAGAGCGCTGTGTGGAGCAAGATAAGATATTTCAGGAAGAAA 864
DB 354 LysTrpLysThrTyrlsGlyIleValAsnAsnGluLysValPheGlnGlyAsnSer 373
QY 865 GATTATCACAGGATGTGCGGTAACTATTTTGGCACCAGTATTTGACAGCTTTTATAGA 924
DB 374 AsnPheArgAspProValGlnAsnAsnPheIleProProIleValAlaArgTyrlsValArg 393

QY 925 GTGAATCTACCCCAATGGCAGCAGAAATTTGCCATGAAATGGAGTCTGCGATCTCAG 984
DB 394 ValValProGlnThrTrpHisGlnArgIleAlaLeuLysValGluLeuIleGlyCysGln 413
QY 985 TTTATTCCTAAAGGTCCTCCAAACACTTACTCAACCTCCACCTCCCTCGGACAGCAAT 1044
DB 414 -----TleThrGln-----GlyAsn 418
QY 1045 GAC-----CTCAAAACACTACAGCCCTCCCAAAATAGCCAAAGCTCGTCC 1092
DB 419 AspSerLeuValTrpArgLysThrSerGlnSerThrSerValSerThrLysLysGluAsp 438
QY 1093 CCAAAATTTACCAACCACTACCACTCGCAGTAGCAATGAATTTCTCCACAGACAGAA 1152
DB 439 GluThrIleThrArgProIle-----ProSerGlu 448
QY 1153 CAACAACCTGCCAGTCTGATATCAGAAATATACCGTAACTCCAAATGTAAACCAAGAT 1212
DB 449 GluThr-----SerThrGlyIleAsnIleThrThr----- 458
QY 1213 GTAGCGCTGGCTGCAGTCTTCTCCTGCTGCTGCTCATGCTCCTACTCTCTCTCTC 1272
DB 459 ValAlaIleProLeuValLeuValValLeuValPheAlaGlyMetGlyIlePheAla 478
QY 1273 ATATTAGTGTGTCTGGCACTGGAGAAACAGAAAGAAAGAAAGTGAAGCACCTATGAC 1332
DB 479 AlaPhe-----ArgLysLysLysLysLysGlySer----- 488
QY 1333 TTACCTTACTGG-----GACCGGCGAGTGTGGTGAAGGAATGAAGACAGTTT 1380
DB 489 ---ProTyrglySerAlaGluAlaGlnLysThrAspCysTrpLysGlnIleLysTy--- 506
QY 1381 CTTCTCGCAAAAGCAGTGGACCATGAGAAACCCCA---GTTCCGTATACAGCAGCGAA 1437
DB 507 -----ProPheAlaArgHisGlnSerAlaGlu 515
QY 1438 -----GTTAATCACCTGAGTCCAGAGAGTCAAC-----ACAGTGTGCGAGGT 1482
DB 516 PheThrIleSerTyraAspAsnGluLysGluMetThrGlnLysLeuAspLeuIleThrSer 535
QY 1483 GACTCTGCAGATATCTCAGCCACTGGTAGGAGGAGTGTGGTACACATTCATCAAGA 1542
DB 536 AspMetAlaAspTyrglnGlnProLeuMetIleGlyThrGlyThrValThrArgLysGly 555
QY 1543 TCTACCTTTAAACCA-----GAAGAAGGAAAGAGCAGGCTATGACAGACCTAGATCCT 1596
DB 556 SerThrPheArgProMetAspThrAspAlaGluGluAlaGly---ValSerThrAspAla 574
QY 1597 -----TACAACCTACACAGGCGAGAGTATTATCCTCTATGCTGAACCACTCCCA 1647
DB 575 GlyGlyHisTyraAspCysProGlnArgAlaGlyArgHisGluTyraLeuProLeuAla 594
QY 1648 ATTACGGGCGCTGATGATCAACCCCAATCATCATGACATGTCAGGCGACCCCACT 1707
DB 595 ProProGluProGluTyraThrProIleVal-----GluArgHisValLeuArg 611
QY 1708 TCAGTTGGTGCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACT----- 1758
DB 612 AlaHisThrPheSerAlaGlnSerGlyTyraArgValProGlnProGlnProGlyHisLys 631
QY 1759 -----CCCCACTA-----GTGGGAACCTTCAATACACTT 1788
DB 632 HisSerLeuSerSerGlyGlyPheSerProValAlaGlyValGlyAlaGlnAspGlyAsp 651
QY 1789 CTTCTCAGAGACTCAGAGCTCTCTCAGCCAGGCGCCAGTATGATACCCGAAAGCT 1845
DB 652 TyrGlnArgProHisSerAlaGlnProAlaAspArgGlyTyraSerProLysAla 670

RESULT 9
AAU00628
ID AAU00628 standard; Protein: 487 AA.
XX

AAB19126 standard; Protein; 503 AA.
 AAB19126;
 19-FEB-2001 (first entry)
 Polypeptide isolated from lymph node stromal cells of fsn +/- mice.
 Lymph node stromal cell; fsn +/- mice; inflammatory disorder;
 immune system disorder; cancer; viral infection; HIV infection;
 blood vessel growth; tumour necrosis factor disorder; arthritis;
 inflammatory bowel disease; fibroblast growth factor-mediated disorder;
 cardiac failure.
 Mus sp.
 WO200058463-A1.
 05-OCT-2000.
 18-FEB-2000; 2000WO-NZ00015.
 25-MAR-1999; 99US-0276268.
 26-AUG-1999; 99US-0383586.
 (GENE-) GENESIS RES & DEV CORP LTD.
 Strachan L, Sleeman M, Abernethy N, Orrust R, Kumble KD;
 Murison JG;
 WPI: 2000-664924/64.
 N-PSDB; AAA96736.
 Polypeptide expressed in mammalian fsn +/- lymph node stromal cells,
 useful for modulating growth of blood cells, for treating inflammatory
 and tumour necrosis factor-mediated disorders, cancer and viral
 disorders
 Claim 1; Page 68-69; 75pp; English.
 The present sequence represents a polypeptide sequence which is
 isolated from lymph node stromal cells of fsn +/- mice. The
 polynucleotides and their polypeptides are useful for treating an
 inflammatory disorder, disorder of immune system and cancer selected
 from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
 viral disorder, in particular HIV infection and for modulating the
 growth of blood vessels. The polypeptides are useful for treating a
 tumour necrosis factor (TNF) mediated disorder, such as those selected
 from arthritis, inflammatory bowel disease and cardiac failure and a
 fibroblast growth factor-mediated disorder. It is also useful in assays
 to determine biological activity, to raise antibodies, to isolate
 corresponding ligands or receptors, to quantify levels of protein or
 cognate corresponding ligand or receptors, as anti-inflammatory agents,
 and in compositions for the treatment of skin, connective tissue and
 immune system diseases. The polynucleotide is useful as marker for
 tissue, as a chromosome marker or tags in the identification of a
 genetic disorder.
 SQ Sequence 503 AA;
 Alignment Scores:
 Pred. No.: 2,87e-29 Length: 503
 Score: 416.00 Matches: 148
 Percent Similarity: 32.98% Conservative: 70
 Best Local Similarity: 22.39% Mismatches: 147
 Query Match: 11.63% Indels: 296
 DB: 21 Gaps: 19
 US-10-060-830-1113 (1-1962) x AAB19126 (1-503)
 QY 37 CTGCTGCTCGAGACCTGGAGCCCAAGCAAAATACCTGTGTGGTTCGAAATG 96
 Db 90 LeuLeuPheSerSerAlaThrAspGlnThrGlyProTyrCysGly---SerTrpAlaVal 108
 AAB19126

QY 97 AACCAATTCATTAATGAAACCAAGCAATGAAATCACATTCGCTGTTTCATGAGTGAATCCAT 156
Db 109 ProLysGluLeuArgLeuAsnSerAsnGluValThrValLeuPheLysSerGlySerHis 128
QY 157 GTTCTGGACGGGATTTTGGCTCATCTCTGTTATAGATAAACAAGACTCAATATCT 216
Db 129 IleSerGlyArgPheLeuLeuThrTyAlaSerSerAspHisProAspLeuIleThr 148
QY 217 TGTGTGGACACTGCATCCAAATTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGGT 276
Db 149 CysLeuGluArgGlySerHisTyPheGluGluLysTySerLysPheCysProAlaGly 168
QY 277 TGTCTCTCTCTTCTCTGATGATCTGGAACAATTCCTCATGATAGATGATTCCTCG 336
Db 169 CysArgAspIleAlaArgAspIleSerGlyAsnThrLysAspGlyTyArgAspThrSer 188
QY 337 CCATTGTGATGCTGCTGTCATGAGGAGTAGTGTCAAAACAGTGTGGCGGCCAAATC 396
Db 189 LeuLeuCysLysAlaAlaIleHisAlaGlyIleIleThrAspGluLeuGlyGlyHisIle 208
QY 397 AGTGTGTAATAGTAAAGTATTCCTATTATGTAAGTCTTTGGCTTAACAACGTCACA 456
Db 209 AsnLeuLeuGlnSerLysGlyIleSerHisTyGluGlyLeuLeuAlaAsnGlyValLeu 228
QY 457 TCTGTGTGGACACTTACTACAAGTCTTTTACATTTTAAAGACAAGTGGATGTATGGA 516
Db 229 SerArgHisGlySerLeuSerGluLysArgPheLeuPheThrThr 243
QY 517 ACATGGGATGGAGTCTGTGTGATCGGGGATCCTCAAATAACAGCATCATCTGTGCTG 576
Db 243 243 243
QY 577 GAGTGACTGACACACAGGGCAAGAGACAGTGTGGAACCCAAAAGCCAGCTGAAA 636
Db 243 243
QY 637 AAACCTGGACCGCTTGGCTGCTTTTGCCTGATGATACCATGCTTACAAATAGAT 696
Db 243 243
QY 697 TTGAATAAGGAAAAGAAAATAACAGCATTATAACCACTGGATCCACCATGGTGGACAC 756
Db 243 243
QY 757 AATTACTATGTCTGCTACAGAAATCCTGTACAGTATGATGGGCAAGAAATGGACTGTG 816
Db 243 243
QY 817 TACAGAGAGCCTGTGTGGAGCAAGATAAGATATTTCAAGGAAACAAAGATTATCACCAG 876
Db 244 244
QY 877 GATGTGCGTAATAACTTTTGGCCACCAATATTGACAGTATTTATTAGAGTGAATCCPACC 936
Db 246 246
QY 937 CAATGGCAGCAGAAATGCCATGAAATGGAGCTGCTCGGATGTCAGTTTATTCCTAAA 996
Db 246 246
QY 997 GGTGCTCTCAAACTTACTCAACCTCCACCTCGGAACAGCAATGACCTCAAAAAC 1056
Db 246 246
QY 1057 ACTACAGCCCTCCAAAATAGCCAAAGTGTGCCCCAAAATTTAGCAACCACTACAA 1116
Db 246 246
QY 1117 CCTCGCATGAGCAATGAATTTCTGCACAGACAGACAAACAACACTGCCGCTCTGTATC 1176
Db 246 246

QY 1177 AGAAATACTACCGTAACCTCCAAATGTAACCAAGATGTAGCGCTGCTGCTGCTTCTGTC 1236
Db 247 247
QY 1237 CTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1296
Db 258 258
QY 1297 AGAAACAGAAAGAAAACAACTGAAGCAGCCTATGACTTACCTTAC 1344
Db 273 273
QY 1345 GACCGGCGAGTGTGGAAAGAAATGAAGCAGTCTTCTCTGCAAAAGCAGTGGACCAT 1404
Db 288 GlnLysThrGlyCysTrpLysGlnIleLysTy 298
QY 1405 GAGGAACCCCA---GTTGCTATAGCAGCAGGAA---GTTAATCACCTGAGTCCA 1455
Db 299 299
QY 1456 AGAAGAGTCAAC---ACAGTCTGCTGAGGCTGACTCTGCAGAGTATGCTCAGCCA 1506
Db 316 LysGluMetThrGlnLysLeuAspLeuIleThrSerAspMetAlaAspTyGlnGlnPro 335
QY 1507 CTGCTAGGAGGAATTTGTTGCTACATTCATCAAGATCTACCTTTAAACCA--- 1557
Db 336 LeuMetIleGlyThrGlyThrValAlaArgLysGlySerThrPheArgPrometAspThr 355
QY 1558 1558
Db 356 AspThrGluGluValArgValAsnThrGluAlaSerGlyHis 370
QY 1600 AACTCAGCAGGCGAGGAAGTTTATCATGCTGCTGCAACCACTCCCAATTCAGGGCCT 1659
Db 371 AspCysProHisArgProGlyArgHisGluTyAlaLeuProLeuThrHisSerGluPro 390
QY 1660 GAGTATGCAACCCCAATCATC--- 1689
Db 391 GlutylAlaThrProIleValGluArgHisLeuLeuArgAlaHisThrPheSerThrGln 410
QY 1690 TCAGGCAACCCCAACACTTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1749
Db 411 SerGlyTyArgValProGly---ProArgProThrHisGluHisSerHisSerSerGly 429
QY 1750 AACCAACCTCCCTCAGTGTGGGAACTTACATACACTTCTCTCCAGGACTGACAGCTGC 1809
Db 430 GlyPheProProAlaThrGlyAlaThrGlnValGluSerTyThrGlnArgProAlaSerPro 449
QY 1810 TCCTCAGCCCGCCCATGATATACCCCGAAAGCTGGGAGCCAGCTCTACCTGCCCA 1869
Db 450 LysProValGlyGlyTyArgLysPro---AlaAlaSerSerPheLeuAspSerArg 468
QY 1870 GAGCAATGTGTACAGGTCGCCAGAGCAGACACAGAGATATCAGGAGCAGGAGGAT 1929
Db 469 Asp 483
QY 1930 GGG 1932
Db 483 Gly 483
RESULT 13
AAE22715
ID AAE22715 standard; Protein; 398 AA.
XX AAE22715;
AC XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hyl protein.
XX Human; neuropilin-like polypeptide; neuropilin-Hyl; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;

Db 379 LysGlyIleValAsnAsnGluGluLysVal---ArgGlyAsn 391

RESULT 14
AAU79459
ID AAU79459 standard; Protein; 398 AA.
XX AC AAU79459;
XX DT 15-JUL-2002 (first entry)
XX XX Human Neuropilin-Hy1.
XX DE
XX KW Human; neuropilin-Hy1; chromosome 6q21; neuronal growth;
KW nerve regeneration; neurodegenerative disease; learning disorder;
KW memory disorder; Alzheimer's disease; angiogenesis; neovascularisation;
KW organ growth; nervous system lesion; cancer; cell proliferation;
KW cell differentiation; stem cell growth factor activity;
KW Parkinson's disease; Huntington's disease; amytrophic lateral sclerosis;
KW haematopoiesis; myeloid cell disorder; lymphoid cell disorders; anaemia;
KW platelet disorder; thrombocytopaenia; liver fibrosis; lung fibrosis;
KW reperfusion; food supplement; DNA microarray.
XX KW
XX OS Homo sapiens.
XX QS
XX PN WO200222780-A2.
XX XX
XX PD 21-MAR-2002.
XX XX
XX PF 11-SEP-2001; 2001WO-US28590.
XX PF
XX PR 11-SEP-2000; 2000US-0659671.
XX PR 06-SEP-2001; 2001US-0659671.
XX XX
XX PA (TANG/) TANG T Y.
XX XX
XX PI Tang TY;
XX XX
XX WP1; 2002-351881/38.
XX DR N-PSDB; ABK49565.
XX XX
XX PT New neuropilin-like polypeptides for diagnosing, preventing and
PT treating neurological conditions and disorders, cancers, and for
PT inducing angiogenesis and neovascularisation -
XX XX
XX PS Claim 3; Page 118-120; 144pp; English.
XX XX
XX CC The invention relates to an isolated neuropilin-like polypeptide
CC including neuropilin-Hy1 and neuropilin-Hy2, the full length cDNAs
CC encoding the proteins and the coding regions of the cDNAs. Also included
CC is a nucleic acid array comprising the cDNAs attached to a surface used
CC for detecting full-matches or mismatches to the cDNAs. The genes
CC for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21.
CC The nucleic acid array is useful for detecting full-matches or mismatches
CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
CC in modulating neuronal growth, regenerative capacity, treating
CC neurodegenerative diseases, learning and memory disorders, diagnosing and
CC mapping genetic neuronal defects and degenerative diseases like
CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
CC and organ growth and development (e.g. the heart). The nervous system
CC disorders include lesions of central or peripheral nervous systems,
CC including traumatic lesions, ischaemic lesions, infectious lesions,
CC degenerative lesions, lesions associated with nutritional diseases or
CC disorders, neurological lesions, and lesions caused by toxic substances.
CC The neuropilin-like proteins and cDNAs are also useful as markers for
CC cancers. The neuropilin-like proteins are useful for regulating cell
CC proliferation, cell differentiation, stem cell growth factor activity,
CC for inducing proliferation of neural cells, regeneration of nerve and
CC brain tissue, for treatment of central and peripheral nervous system
CC diseases, and neuropathies, such as Parkinson's disease,
CC Huntington's disease, amytrophic lateral sclerosis, to regulate
CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
CC anaemias, and platelet disorders, such as thrombocytopaenia,
CC regeneration and treatment of lung or liver fibrosis, reperfusion

CC	injury in various tissues and as a food supplement or molecular weight marker. The cDNAs are useful in gene identification, genome mapping, transgenics, as hybridisation probes, for primer design, for gene chips and as a DNA antigen. The present sequence represents neuroligin-Hy1.
XX	Sequence 398 AA;
50	Alignment Scores:
	Pred. No.: 8,08e-26 Length: 398
	Score: 378.00 Matches: 104
	Percent Similarity: 42.09% Conservative: 45
	Best Local Similarity: 29.38% Mismatches: 99
	Query Match: 10.62% Indels: 106
	DB: 23 Gaps: 9
US	10-060-830-1113 (1-1962) x AAU79459 (1-398)
QY	37 CTGCTGTCGAGGAGCTGGAGCCACGACGCAAAATACGTGCTGTGGGTGCAAAATG 96
DB	65 LeuLeuPheThrSerSerSerAspGlnTrpGlyProTyrCysGly---SerMetThrVal 83
QY	97 AACCATTCAAATGAATCAAAAGCAATGAATACATCTGCTTTTCATGAGTGAATCCAT 156
DB	84 ProLysGluLeuLeuLeuLeuAsnThrSerGluValThrValargPheGluSerGlySerHis 103
QY	157 GTTCTTGACGCGGATTTTTCGCTCATACATCTCTGTTATAGAT----- 198
DB	104 IleSerGlyArgGlyPheLeuLeuThrTyrAlaSerSerAspHisProGluSerGlnGly 123
QY	199 -----AAACAGATCTAATTACTTGTGTTGGACATGCATCC 234
DB	124 AspArgProSerGluLysThrLeuAspGlnGlnSerArgThrPheLeuAlaThrGlyThr 142
QY	235 RAATTTTGGAACTGAGTCTCAGTAAGTACTGCCAGCTGCTGCTCTCTCTTTTGCT 299
DB	144 ThrPheValLysAspSerPheSer----- 151
QY	295 GAGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTGCCATGTGCATGGCTGGT 350
DB	152 ---ThrAspGlyThr-----SerLeuLeuCysLysAlaAla 163
QY	355 GTGCATGCAGGAGTAGTCTCAACACAGCTTGGCGGCCAAATCAGTGTGTAAATAGTAA 410
DB	163 IleHisAlaGlyIleIleAlaAspGluLeuGlyGlnIleSerValLeuGlnArgLys 180
QY	415 GGTATTCCTATTATGAAGTCTTGTGCTTAACACGTCACATCT----- 450
DB	183 GlyIleSerArgTyrGluGlyIleLeuAlaAsnGlyValLeuSerArgGluPheGluIle 201
QY	459 ----- 459
DB	203 PheArgGluGlnLeuPheSerSerValLeuPheTyrSerTrpGlyAsnThrValHisAla 222
QY	460 -----GTGGTGGGACAC----- 470
DB	223 ValIleGluLeuMetPheProHisMetIleValTrpHisSerGlyLysThrArgGluGly 241
QY	471 ----- 471
DB	243 SerIleAlaAlaGluGluGlyValProLysLeuTyrLeuValIleGlnLysGlnGlu 261
QY	472 TTATCTACAAGTCTTTTACATTTAAGACACAGTGGATGTTATGGAACACTGGGATGGAG 533
DB	263 LeuValGlnAspLeuValLeuValAlaThrValGlyCysSerArgSerLeuSerPheGlu 280
QY	532 TCTGTGTGATCCGGATCCTCAAAATAACAGCATCATCTGTCTGTGGATGGCATGCCAC 591
DB	283 -----ProAspGlyGlnIleArgAlaSerSerSerTrpGlnSerValAsnGlu 299
QY	592 ACAGGCCAAGACACAGTTGGAAACCCAAAGCCAGCGCTGAAACACCTGGACCGCT 651
DB	299 SerGlyAspGlnValHisTrpSerProGlyGlnIleArgLeuGlnAspGlnGlyProSer 310

RESULT 15

AW383889 664 bp mRNA linear EST 04-FEB-2000
LOCUS QV3-HT0364-241199-024-b01 HT0364 Homo sapiens cDNA, mRNA sequence,
DEFINITION AW383889
ACCESSION AW383889
VERSION AW383889.1 GI:6988548
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 664)

CGCP <http://www.ludwig.org.br/ORESTES>.

AUTHORS The FAPESP/LICR Human Cancer Genome Project

TITLE Unpublished (1999)

JOURNAL Contact: Simpson A.J.G.

COMMENT Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3&t2=QV3-HT0364-241199-024-b01&t3=1999-11-24&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 53

High quality sequence stop: 623.

FEATURES

source

1..664
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0364"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 185 a 130 c 157 g 192 t
ORIGIN

Alignment Scores:

Pred. No.:	2,28e-91	Length:	664
Score:	1011.00	Matches:	193
Percent Similarity:	96.53%	Conservative:	2
Best Local Similarity:	95.54%	Mismatches:	7
Query Match:	26.00%	Indels:	0
DB:	10	Gaps:	0

US-10-060-830-3 (1-729) x AW383889 (1-664)

QY	112	IlleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGly	131
Db	53	ATTGAATCAAAAGCAATGAATCAACATAGCTGTTGAGTGAATCCATGTTCTGGA	112
QY	132	ArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAsp	151
Db	113	CGCGGATTTTGGCCCTCATCTCTGTATAGATAAACAAGATCTAATATTGTTGGAC	172
QY	152	ThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeu	171
Db	173	ACTGCATCCAAATTTTGGGAACCTGAGTTCAGTAAGTACTGCCAGCTGTTGCTGCTT	232
QY	172	ProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerSerProLeuCys	191
Db	233	CCTTTTCTGAGATATCTGGAACAAATTCCTCATGGATAGAGATTCTCGCCATTGTGC	292
QY	192	MetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlnIleSerValVal	211

Db	293	ATGGCTGGTGTGCATGTCAGGAGTAGTGTCAAACAGCTTGGCGGCCAAATCAGTGTGTGA	352
QY	212	IlleSerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThrSerValVal	231
Db	353	ATTAGTAAAGGTATTCCCTATTATGAAAGTTCTTTGGCTAACACAGTCACATCTGTGGTG	412
QY	232	GlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGly	251
Db	413	GCACACTTATCTACAGTCTTTTACATTTAGACAAGTGGATGTTATGGAACACTGGGG	472
QY	252	MetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluIleThr	271
Db	473	ATGGAGTCTGGTGTGATCGCGGATCCTCAATAACAGCATCATCTGTGCTGGAGTGGACT	532
QY	272	AspHisThrGlyGlnGluAsnSerTyrLysProLysLysAlaArgLeuLysLysProGly	291
Db	533	GACCACACAGGCCCAAGAGAACAGTTGGAAACCCCAAAAGCCAGGCTGAAAAACCTGGA	592
QY	292	ProProTyrAlaAlaPheAlaThrAspGluTyrGlnTyrLeuGlnIleAspLeuAsnLys	311
Db	593	CCGGCTTGGGCTGCTTTTGCACCTGATGATACCAGTGGTTCCCATAGAAATTTGATTACG	652
QY	312	GluLys	313
Db	653	GAAAGA	658

Search completed: January 23, 2003, 15:16:34

Job time : 2598.12 secs

QY 194 GlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSer----- 209
Db 1111 AAACCCAGGATGAGATCATCCCTACAGTTCTGACCTTTGGAGCAGTACCCCTCTA 1170
QY 210 -----ValValIleSerLysGlyIlePro--- 217
Db 1171 CAAGTGGGGAAGAGACTGTAATATGACTGGCTGGACATCTGGGATGGCATTCACAT 1230
QY 218 -----TyrTyrGluSerSerLeuAlaAsnAsnValThrSerVal 230
Db 1231 GTTGGACCTCTGATTGGCAAGTACTGTGGACGAAACACCCCTCCAACTCCGCTCGTC 1290
QY 231 ValGlyHisLeuSerThrSerLeuPheThrPheLysThrSer----- 244
Db 1291 ACGGGATCTCTCC-----TTGACCTTTACACAGGACATGGCAGTGGCCAAAGGAT 1341
QY 245 -----GlyCysTyr 247
Db 1342 GGCTTCTCCGACGTACTATTGATCCACAGAGCCACTGAGAATTTTCAGTGCAT 1401
QY 248 GlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerVal 267
Db 1402 GTCCCTTTGGGAATGGAGTCTGCCGGATTGCTAATGAACAGATCAGTGCCTCTCC- 1458
QY 268 LeuGluThrPheThrAspHisThrGlyGlnGluAsnSerTrpLysLysAlaArgLeu 287
Db 1459 -----ACCTTCTGATGGGAGTGGACTCCTCAACAGAGCGGCTC 1500
QY 288 LysLysProGlyProProTyrAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIle 307
Db 1501 CATGGTGATGACAAAGGCTGGACACCCAAATTTGGATTCCAAAGGAGTATCTCCAGGTG 1560
QY 308 AspLeuAsnLysGluLysLysIleThrGlyIleIleThrGly-----SerThrMet 325
Db 1561 GACCTGCGCTCTCAACCATGCTCACAGCCATTGCAACACAGAGGCCATTTCACGGGAA 1620
QY 326 ValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspAspGlyGlnLys 345
Db 1621 ACCCAAGAAAGCTACTACGCTCAATATGCTAAGCTGGAAGTCAGCACAATGGTGAAGAT 1680
QY 346 TrpThrValTyrAspGluProGlyValGlnAspLysIlePheGlnGlyAsnLysAsp 365
Db 1681 TGGATGTCTACCGGCATGGC-----AAACACACAGATATTCACAGCAGCAATGAT 1734
QY 366 TyrHisGlnAspValArgAsnAspPheLeuProIleIleAlaArgPheIleArgVal 385
Db 1735 GCGACCGAGGTGGTCTAAACAAGCTCCACATGCCACTGCTGACTCGGTTTCATCAGGATC 1794
QY 386 AsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuGlyCysGlnPhe 405
Db 1795 CGCCCGCAGACGTGGCATTTGGGCATTTGGCCTTCCCTCGCTGGAGCTTTTGGCTCGCGGTC 1854
QY 406 IlePro-LysGlyArgPro-----ProLys 413
Db 1855 ACAGATGACCCCTGCTCAACATGCTGGGGATGCTCTCGGGCCTCATTTGCTGATACCCAG 1914
QY 413 sLeuThrGlnProProProProArgAsnSerAsn----- 424
Db 1915 ATCTCTGCTCTCCACCCGAGAGTACCTCTGGAGCCCGCAGTGTGCGCTGGTTAGT 1974
QY 425 -----AspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAl 440
Db 1975 AGCCGCTCTGGCTGTTTCTCGGAACCCCTCAAGCCCGCCAGGAGTGAAGATGGCTTCAG 2034
QY 440 aProLysPheThrGlnProLysProGlnProArgSerSerAsnGluPheProAlaGlnThrGln 460
Db 2035 GTTGACCTGGGACACCCCAAGACATGAAAGGGGTATCATCTCC-----AGGGAGCCCGA 2088
QY 460 uGlnThrThrAlaSerProAspIleArg----- 469
Db 2089 GGAGGACAGCATCAGTCCCTGGAGACCGCGGCTTTGTACGCAAGTTCAAAGTCTCC 2148
QY 470 -----AsnThrThrValThrProAsnValThrLysAspVa 481

Db 2149 TACAGCTAAATGGCAAGGACTGGGAATATATCCAGACCCCGAGGACTCAGCAGACAAG 2208
QY 481 lAlaLeu-----AlaAlaValLeuValProValLeuValMetValLeuThrThrLeuIle 499
Db 2209 CTGTTTGAAGGGAACATGCATATGACACCC-----CTGACATCCGGAAG 2253
QY 499 eLeuIleLeu-----ValCysAlaTrpHisTrpArgAsnArgLysLysLysTh 515
Db 2254 TTGATCTCTTCCAGCGCAGTATGTGCGGGTGTACCCAGAGAGGTGGTCCGCCAGCAGC 2313
QY 515 rGlu-----GlyThrTyrAspLeuProTyrTrpAs 525
Db 2314 ATCGGATGAGGCTGGAGTGTGGCTGTGACTGACAGACTCAAGCCACAGTGGAG 2373
QY 525 pArgAlaGlyTrpTrp-----LysGlyMetLysGlnPheLeuProAl 539
Db 2374 ACGC-----TGGGACCCACCGTGAAGAGTGAAGAGACTACCAACCCATATCCCATG 2424
QY 539 aLysAlaValAspHisGlu-----GluThrProValArgTyrSerSerSerGluValAs 557
Db 2425 GATGAGGATGCCACCGAGTGTGGGAAACTGCAGCTTTGAGGATGACAAGATTGCAA 2484
QY 557 nHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyrAlaGln 577
Db 2485 CTTCCTTCAGATTCAACTGCACTTTGATTTCGG-----AAGAG 2526
QY 577 nProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGln 597
Db 2527 ACCTGTGTTGG-----TGTACGACCATGCCAAGTGG 2559
QY 597 uGlyLysGluAlaGlyTyrAlaAspLeuAspProTyr-----Asn-SerProGlyGlnG 615
Db 2560 CTCGGAGCAGCTGGATCAGCAGCGCTAACCCCAATGACAGAACATTTCCAGATGACAAG 2619
QY 615 luValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThr----- 633
Db 2620 AACTCTTGAACI---GCAGATGATGGCGGAGGCGGAGGCGGAGTACGGGCGCTCATC 2676
QY 634 -----ProIleIleMetAspMetSerGlyHisProThr 645
Db 2677 AGCCACCGGTGACCTGCCCCGAGCCCTGTGTG-----CATGGAGTTCCAGTAC 2727
QY 645 hrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProProL 665
Db 2728 CAAGCCATGGCGCGCAGCGGGTGGCTGAGTGTTCGGGAAGCCAGCA----- 2780
QY 665 euValGlyThrTyrAsnThrLeu-LeuSerArgThrAspSerCysSerSerAlaGlnAla 684
Db 2781 -----GGAAACCAACTCTTTGGGTCTATCGTGAGGACCGGCGAGTGG----- 2829
QY 685 GlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTyr 704
Db 2830 -----AAGCACGGCGCATTTATCTGCCAGCTATGAC---ATGGAGTAT 2871
QY 705 GlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGlyGluCysAspVal 724
Db 2872 CAGATC-----GTGTTGAGGGAGTGTATAGGGAAGGAGCAGTCCGGAGAGATTTCATC 2925

RESULT 3

US-08-936-135-21

; Sequence 21, Application US/08936135

; Patent No. 6054293

; GENERAL INFORMATION:

; APPLICANT: Tessier-Lavigne, Marc

; APPLICANT: He, Zhigang

; APPLICANT: Chen, Hang

; TITLE OF INVENTION: Semaphorin Receptors

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 75 DENISE DRIVE

; CITY: HILLSBOROUGH

Db 969 TTCCTCTACGCTATGAGATCTTCAAAACAGGCTCT-----GAAGATTGTTC 1016
 Qy 154 SerAsnProGluProGluPheSerLysThrCysProAlaGlyCysLeuLeuProPhe 173
 Db 1017 AAGAACTTTACAAGCCCAATGGGACCAATGAATCTCCAGG-----TTT 1061
 Qy 174 AlaGluLeuSerGlyThrIleProHisGlyThrArgAspSerSerProLeuCysMetAla 193
 Db 1062 CCAGAG-----AAGATCCACACAATCTG---GACTGTACCTTCACCATCTCGGCC 1109
 Qy 194 GlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSer----- 209
 Db 1110 AAACCCAGGATGGAGATCATCTACAGTTCTCTGACCTTTGACCTTGGAGCATGACCTCTA 1169
 Qy 210 -----ValValIleSerLysGlyIlePro--- 217
 Db 1170 CAAGTGGGGAAGAGAGCTGTAATATGACTGGCTGGACATCTGGGATGGCATTCACAT 1229
 Qy 218 -----TyrTyrGluSerSerLeuAlaAsnValThrSerVal 230
 Db 1230 GTTGGACCTCTGATTGGCAAGTACTGTGGGACGAAACACCTCCAACTCCGCTCGTCC 1289
 Qy 231 ValGlyHisLeuSerThrSerLeuPheThrPheLysThrSer----- 244
 Db 1290 ACGGGATCCTCTCC-----TTGACCTTTACACGGACATGGCAGTGCCCAAGAT 1340
 Qy 245 -----GlyCysTyr 247
 Db 1341 GCGTCTCGGCACGTTACTATTGTATCCACCAGGACACCTGAGAAATTTTCAGTGCAAT 1400
 Qy 248 GlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerVal 267
 Db 1401 GTCCCTTTGGGAATGGAGTCTGCGCGGATGTCTAATGAACAGATCAGTGCCTCTCC--- 1457
 Qy 268 LeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeu 287
 Db 1458 -----ACCTTCTGTATGGGAGGTGGACTCTCTCAACAGACGCGGCTC 1499
 Qy 288 LysLysProGlyProProTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIle 307
 Db 1500 CATGGTGATGACAAATGGCTGGACACCAATTTGGATTCCACAGGAGTATCTCCAGGTG 1559
 Qy 308 AspLeuAsnLysGluLysLysIleThrGlyIleIleThrGly-----SerThrMet 325
 Db 1560 GACCTCGCTTCTTAACCATGTCTACAGCATTTGCAACACAGGAGCCATTTCCAGGAA 1619
 Qy 326 ValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLys 345
 Db 1620 ACCCAAGAGGCTACTACGTCAATCGTCAACAGCTGGAAGTCAGCACAAATGGTGAAGAT 1679
 Qy 346 TrpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAsp 365
 Db 1680 TGGATGGTCTACCGCATGGC-----AAAAACACACAGATATCCAAAGCAACAAATGAT 1733
 Qy 366 TyrHisGlnAspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgVal 385
 Db 1734 GCGACCGAGGTGTGCTAAACAAGCTCCACATGCCACTGCTGACTCGGTTCATCAGATC 1793
 Qy 386 AsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPhe 405
 Db 1794 CGCCCGCAGACGCTGGCATTTGGCATTTGCCCTTCGCTGGAGCTCTTTGGCTGCCGGTTC 1853
 Qy 406 IlePro-LysGlyArgPro-----Proly 413
 Db 1854 ACAGATGACCTGTCTCAACATGCTGGGGATGCTCTCGGGCTCATTTGCTGATACCCAG 1913
 Qy 413 sLeuThrGlnProProProProArgAsnSerAsn----- 424
 Db 1914 ATCTCTGCCTCTCCACCCGAGAGTACCTCTCGAGCCCCAGTGTGCGCGCTGGTTAGT 1973
 Qy 425 -----AspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla 440
 Db 1974 AGCCGCTCTGGCTGGTTCTCTCGGAACCTCAAGCCAGCCAGGTTGAAGATGGCTTCAG 2033

Qy 440 aProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrG1 460
 Db 2034 GTAGACCTGGGAGCACCCACAGACAGTGAAGGGGTTCATCATCC-----AGGGAGCCGA 2087
 Qy 460 uGlnThrThrAlaSerProAspIleArg----- 469
 Db 2088 GGAGGAGACACATCACTGCGGTGGAAGCCAGCGGGTGTGTACGCAAGTTCAAAAGTCTCC 2147
 Qy 470 -----AsnThrThrValThrProAsnValThrLysAspVa 481
 Db 2148 TACAGCCTAATGGCAAGGACTGGGAATATATCCAGGACCCAGGACTCAGCAGACAAG 2207
 Qy 481 lAlaLeu-----AlaAlaValLeuValProValLeuMetValLeuThrLeuI1 499
 Db 2208 CTGTTTGAAGGAACATCACTATGACACCC-----CTGACATCCGAGG 2252
 Qy 499 eLeuIleLeu-----ValCysAlaTrpHisTrpArgAsnArgLysLysLysTh 515
 Db 2253 TTCGATCTCTGTTCCAGCGCAGTATGTGCGGTGTACCCAGAGAGGTGTGCGCAGCAGC 2312
 Qy 515 rGlu-----GlyThrTyrAspLeuProTyrTrpAs 525
 Db 2313 ATCGGGATGAGCTGGAGGTGCTGGGCTGTGACTGGACAGACTCAAGCCACAGTGGAG 2372
 Qy 525 pArgAlaGlyTrpTrp-----LysGlyMetLysGlnPheLeuProAl 539
 Db 2373 ACGC-----TGGGACCCACCGTGAAGAGTGAAGAGACTACCCACCATATCCCATG 2423
 Qy 539 aLysAlaValAspHisGlu-----GluThrProValArgTyrSerSerGluValas 557
 Db 2424 GATGAGGATGCCACCGAGTGTGGGAAAACCTGCAGCTTTGAGGATGACAAAGATTGCAA 2483
 Qy 557 nHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyrAlaGl 577
 Db 2484 CTTCCTTCAGGATTAACCTGCAACTTGTATTTCCGG-----AAGAG 2525
 Qy 577 nProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProThrT 597
 Db 2526 ACCTGTGGTTGG-----TGTACGACCATCCCAAGTGG 2558
 Qy 597 uGlyLysGluAlaGlyTyrAlaAspLeuAspProTyr-----Asn-SerProGlyGlnG 615
 Db 2559 CTCGGGAGCAGTGGATCAGCAGCGCTAACCCCAATGACAGAACAATTTCCAGATGACAAG 2618
 Qy 615 luValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThr---- 633
 Db 2619 AACTTCTTGAACCT---GCAGAGTGTATGCCCGCAGAGAGGCGCATACGGCGGCTCATC 2675
 Qy 634 -----ProIleIleMetAspMetSerGlyHisProThrT 645
 Db 2676 AGCCACCGGTGCACCTGCCCGGAAGCCCTGTGTG-----CATGGAGTTCAGTAC 2726
 Qy 645 hrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProProL 665
 Db 2727 CAAGCATATGGCGGCCACGGGTGCGAGTGTGCTGAGGTGTTTCGGAAGCCAGCCA----- 2779
 Qy 665 euValGlyThrTyrAsnThrLeu-LeuSerArgThrAspSerCysSerSerAlaGlnAla 684
 Db 2780 -----GGAAGCAAACTCCTTTGGTTCATCTGCTGAGGACGAGGCGAGGTGG----- 2828
 Qy 685 GlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTyr 704
 Db 2829 -----AAGACGGGCGCATATTATCTGCCCGCATGTATGAC---ATGAGATAT 2870
 Qy 705 GlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGlyGlyCysAspVal 724
 Db 2871 CAGATC-----GTCTTCGAGGAGTGTATAGGAGGAGGAGGATCGGAGAGATTTCCATC 2924
 RESULT 5
 US-08-936-135-9
 ; Sequence 9, Application US/08936135
 ; Patent No. 6054293

Qy	97	GlulileGlyLysTyrCysGlyLeuGlyLeuGlnMetAsn-----HisSerIleGlu	113
Db	864	CTCCTGGCAAGCACTGGG-----AACATCGCCCGCCCAACATCATC	908
Qy	114	SerLysGlyAsnGlnIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGly	133
Db	909	TCCTCAGGCTCCGTTATACATCAAGTTACCTACGACTACGCCCGCAGGGGGCAGGT	968
Qy	134	PheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAla	153
Db	969	TTCTCTCTACGCTATGAGATCTCTAAACAGGCTCT-----GAAGATTGTTCC	1016
Qy	154	SerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPhe	173
Db	1017	AAGAACATTTACAGCCCAATGGGACCATTAATCTCAGGG-----TTT	1061
Qy	174	AlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerProLeuCysMetAla	193
Db	1062	CCAGAG-----AAGTATCCACACAATCTG---GACTCTACTTCCACCATCTCGGCC	1109
Qy	194	GlyValHisAlaGlyValSerAsnThrLeuGlyGlyGlnIleSer-----	209
Db	1110	AAACCCAGGATGGAGATCATCTACAGTCTCTGACCTTTGACCTGGAGCATGACCTCTA	1169
Qy	210	-----ValValIleSerLysGlyIlePro---	217
Db	1170	CAAGTGGGGGAGGAGACTGTAAATATGACTGGTGGACATCTGGATGGCATCCACAT	1229
Qy	218	-----TyrTyrGluSerSerLeuAlaAsnAsnValThrSerVal	230
Db	1230	GTGGGACCTCTGATWGGCAAGTACTGTGGGACGAAACACCCCTCCAACTCCGCTGCC	1289
Qy	231	ValGlyHisLeuSerThrSerLeuPheThrPheLysThrSer-----	244
Db	1290	ACGGGATCTCTCC-----TTGACCTTCACACGGAGTGGCAGTGGCCCAAGGAT	1340
Qy	245	-----GlyCysTyr	247
Db	1341	GGCTTCTCGCACGTTTACTATTGTATCCACGAGGACCCTGAGAATTTTCAGTGAAT	1400
Qy	248	GlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerVal	267
Db	1401	GTCCCTTTGGGAATGAGTCTGCCGGATTGCTATGAAACAGATCAGTGCCTCTCTCC---	1457
Qy	268	LeuGluTrpThrAspHisThrGlyGlnGlnAsnSerTrpLysProLysLysAlaArgLeu	287
Db	1458	-----ACCTTCTGATGGAGGTGGACTCCTCAACAGACGGCGGCTC	1499
Qy	288	LysLysProGlyProOfTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIle	307
Db	1500	CATGGTGATGACAGTGGCTGGACACCCCAATTTGGATTTCACAACAGGAGATATCTCCAGTG	1559
Qy	308	AspLeuAsnLysGluLysLysIleThrGlyIleIleThrThrGly-----SerThrMet	325
Db	1560	GACCTGGCTTCTTACCATGCTCAGCGCATTTGCAACACAGGAGGCCATTTCCAGGGAA	1619
Qy	326	ValGluHisAsnTyrTrpValSerAlaTyrArgIleLeuTyrSerAspAspGlyGlnLys	345
Db	1620	ACCCAGAAAGCCTACTACGTCAATCGTACAGCTGGAAGTCAGGCACACAATGGTCAAGAT	1679
Qy	346	TrpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnLysAsnLysAsp	365
Db	1680	TGGATGTCTACCGGCATGCC-----AAAAACCAAGATATTTCCAAGCGAACAATGAT	1733
Qy	366	TyrHisGlnAspValArgAsnAsnPheLeuProPheIleAlaArgPheIleArgVal	385
Db	1734	CGCAGCAGAGTGTGCTTAACAAGACTCCACATGCCACTGCTCACTCGGTTTCATCAGGATC	1793
Qy	386	AsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPhe	405
Db	1794	CGCCCGCAGACGTGGCATTTGGCCCTTCGCCTGGAGACTCTTGGCTGCGCGCGGTC	1853
Qy	406	IlePro-LysGlyArgPro-----ProLys	413

Db	1854	ACAGATGCACCTGTCTCCAAACATGCTGGGGANGCTCTCGGGCCCTCATGTGCTGATACCCAG	1913
Qy	413	:LeuThrGlnProProArgAsnSerAsn-	424
Db	1914	ATCTCTGCCTCTCCACCCGAGAGTACCTCTGGAGCCCACTGCTGCCCTGGTTAGT	1973
Qy	425	-----AspLeuLysAsnThrAlaProProLysIleAlaLysGlyArgAl	440
Db	1974	AGCCGCTCTGGTGTCTCTCGGAACCTCAAGCCAGCCAGGATGAAGATGGCTTCAG	2033
Qy	440	aProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGl	460
Db	2034	GTAGACTGGGACCCCAAGACATGAAGGGTCAATCACC-----AGGAGGCCGA	2087
Qy	460	uGlnThrThrAlaSerProAspIleArg-	469
Db	2088	GGAGGAGACAGCATCACTCCCTGGGAGCCAGGGCGTTGTACCAAGTTCAAAGTCTCC	2147
Qy	470	-----AsnThrThrValThrProAsnValThrLysAspVa	481
Db	2148	TACAGCTAAATGGCAAGGACTCGGAATATATCCAGGACCCAGGACTACGACAGCAAG	2207
Qy	481	lAlaLeu-----AlaAlaValLeuValProValLeuValMetValLeuThrThrLeuIl	499
Db	2208	CTGTTTGAAGGGAACATGCATATGACACC-----CTGACATCCCGAAGG	2252
Qy	499	eLeuIleLeu-----ValCysAlaThrHisTrpArgAsnArgLysLysLysTh	515
Db	2253	TTGATCCTGTCTCCAGCGCAGTATGTGGGGTGTACCCAGAGAGGTGTGCGCAGCAGGC	2312
Qy	515	rglu-----GlyThrTyrAspLeuProTyrTrpAs	525
Db	2313	ATCGGATGAGGCTGGAGGTGTGGGCTGTGACTGGACAGACTCAAAAGCCACAGCTGGAG	2372
Qy	525	pArgAlaGlyTrpTrp-----LysGlyMetLysGlnPheLeuProAl	539
Db	2373	ACGC-----TGGGACCCCGTGAAGAGTGAAGAGACTACACCCCATATCCCATG	2423
Qy	539	aLysAlaValAspHisGlu-----GluThrProValArgTyrSerSerGluValas	557
Db	2424	GATGAGGATGCCACCGAGTGTGGGAAACTGCACGTTTGAGGATGACAAAGATTTCGAA	2483
Qy	557	nHisLeuSerProArgGluValThrValLeuGlnAlaAspSerAlaGluTyrAlaGl	577
Db	2484	CTTCTCTCAGGATTCAACTGCAACTTGTATTTCGG-----AAGAG	2525
Qy	577	nProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGl	597
Db	2526	ACCTGTGGTGGG-----TGACGACCATGCCACAGTGG	2558
Qy	597	uGlyLysGluAlaGlyTyrAlaAspLeuAspProTyr-----Asn-SerProGlyGlnG	615
Db	2559	CTCCGGAGCAGCTGGATCAGCAGCGCTACCCCAATGACAGAACATTTCCAGATGACAAG	2618
Qy	615	luValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThr	633
Db	2619	AACCTCTTGAAGTCT--GCAGAGTGTATGCCACGAGAGGGCCAGTACGGCGCTCATC	2675
Qy	634	-----ProIleIleMetAspMetSerGlyHisProThr	645
Db	2676	AGCCACCGGTGCACCTGCCCGAACCCCTGTGTG-----CATGGAGTTCAGTAC	2726
Qy	645	hrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProProProL	665
Db	2727	CAACCCATGGCGGCCACGGGGTGGCAGTCGAGGTGGTTCGGGAAGCCAGCCCA	2779
Qy	665	euValGlyThrTyrAsnThrLeu-LeuSerArgThrAspSerCysSerSerAlaGlnAla	684
Db	2780	-----GGAAAGCAAACCTCTTTGGGTCTCCGTGTAGGACGAGGCGAGTGG	2828
Qy	685	GlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTyr	704

Qy	77	SerAspSerCysHisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThr	96
Db	807	CACGAC---TGCAGATGATGACTTCATTGAGATTGGGATGGGACAGTGAGTCACTGAC	863
Qy	97	GlutIleGlyLysTyrCysGlyLeuGlyLeuGlnMetAsn-----HisIleGlu	113
Db	864	CTCCTGGGCAAGCACTGTGG-----AACATGCCGCCGCCACCATCATC	908
Qy	114	SerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGly	133
Db	909	TCCTCAGGCTCCGTGTATACATCAAGTTCACTCAGACTAGCCCGGCGAGGGCAGGT	968
Qy	134	PheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAla	153
Db	969	TTCTCTACGCTATGAGATCTTCAAAACAGGCTCT-----GAAGATTGTTCC	1016
Qy	154	SerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPhe	173
Db	1017	AAGAACTTTTCAAGCCCAATGGGACCATTTGAATCTCCAGG-----TTT	1061
Qy	174	AlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerSerProLeuCysMetAla	193
Db	1062	CCAGG-----AAGTATCCACACAATCTG---GACTGTACCTTCACCATCTGCCC	1109
Qy	194	GlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSer-----	209
Db	1110	AAACCCAGGATGGAGATCATCTACAGTTCTCCTGACCTTTGACCTGGAGCATGACCTCTA	1169
Qy	210	-----ValIleSerLysGlyIlePro---	217
Db	1170	CAAGTGGGGAGGAGACTGTAATATGACTGGCTGGACATCTGGGATGGCATTCACAT	1229
Qy	218	-----TyrTyrGluSerSerLeuAlaAsnValThrSerVal	230
Db	1230	GTTGGACCTCGATTGGCAAGTAGTCTGGGACCAAAACACCTCCAAACTCCGCTGCTCC	1289
Qy	231	ValGlyHisLeuSerThrSerLeuPheThrPheLysThrSer-----	244
Db	1290	ACGGGATCTCTCC-----TTGACCTTTCACAGGACATGGCAGTGCCCAAGAT	1340
Qy	245	-----GlyCysTyr	247
Db	1341	GGCTTCTCGCACGTTACTATTTTGATCCACGAGGCCACCTGAGAAATTTTCAGTGCAAT	1400
Qy	248	GlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerVal	267
Db	1401	GTCCCTTTGGGAATGGAGTGTGCCGGATTGTGTAATGAACAGATCAGTCCCTCTCC	1457
Qy	268	LeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysIleAlaArgLeu	287
Db	1458	-----ACCTTCTCTGATGGGAGGTGGACTCTCTCAACAGAGCGGGCTC	1499
Qy	288	LysLysProGlyProTrpTrpAlaIlePheAlaThrAspGluTyrGlnTrpLeuGlnIle	307
Db	1500	CATGGTGATGACAATGGCTGGACACCAATTTGGATTTCACAAAGGAGTATCTCCAGTG	1559
Qy	308	AspLeuAsnLysGluLysIleThrGlyIleIleThrThrGly-----SerThrMet	325
Db	1560	GACCTCGGCTTCTTAACCATGTCTCAACGCATTGCAACAGGAGCCATTTCCAGGGAA	1619
Qy	326	ValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLys	345
Db	1620	ACCCAGAAAGGCTACTACGTCAAAATCGTACAAGCTGGGAAGTCAGCAAAATGGTGAAGAT	1679
Qy	346	TrpThrValTyrArgGluProGlyValGlnAspLysIlePheGlnGlyAsnLysAsp	365
Db	1680	TGGATGGTCTACCGGATGGC-----AAAAACACAGATATTCACAGCGAACAATGAT	1733
Qy	366	TyrHisGlnAspValArgAsnAsnPheLeuProIleIleAlaArgPheIleArgVal	385
Db	1734	GCACCGCAGGTGGTGCCTAAACAAAGCTCCACATGCCACTGTGCTGACTCGGTTTCATCAGATC	1793

QY 326 ValGluHisAsnTyrThrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLys 345
Db 1365 ACCCAGAGGGCTACTACGTCATCAATCGTACAGCTGAAGTCAACACACGGGGAAGAT 1424
QY 346 TrpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAsp 365
Db 1425 TGGATGCTCTACCGGATGCG-----AAACACCAAGGTATTCCAGCGCTAACAAATGAT 1478
QY 366 TyrHisGlnAspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgVal 385
Db 1479 GCCACCGAGTGTCTGTGAACAGCTGCACACCGCGCTGTGACTCGTTTCATCAGGATC 1538
QY 386 AsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuGlyCysGlnPhe 405
Db 1539 CGCCCGCAGACGCGGCAATTGGGCATAGCCCTTCGAGCTCTTTGGTTGC----- 1592
QY 406 IleProLysGlyArgProProLysLeuThrGlnProProProProArgAsnSerAsnAsp 425
Db 1593 -----CGGGTCACAGATGCACCGCTCTCCCAACATGCTGGGAATG 1631
QY 426 LeuLysAsnThrThrAlaProLysIleAlaLysGlyArgAlaProLysPhe----- 443
Db 1632 CTCTCGGCGCTCATCTGATACCCAGATCTCTGCCCTCTCCACCGACAGTACCTCTCG 1691
QY 444 -----ThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 460
Db 1692 AGCCCGCAGTGTCCCGCGCTGTGTAGACCGCTCTGCTGCTGCTCCCGAACCCCTCAA 1751
QY 461 GlnThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 480
Db 1752 GCCCAGCCAGGTGAAGTGGCTTACAGTGGATCTTGGGACACCCCAAGACGGTGAAGGC 1811
QY 481 ValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 500
Db 1812 GTCATATCCAGGGGCGCG----- 1831
QY 501 IleLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrAsp 520
Db 1832 -----AGGAGGACAGATATCACTGCCATGGAAGCCAGGCG 1867
QY 521 Leu-----ProTyrTrpAspArgAlaGlyTrpTrpLys 531
Db 1868 ATTTGACGAGTTCGAAGTCTCTACAGCCTTAATGCAAGGACTGGGAATATATCCA 1927
QY 532 GlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThr----- 547
Db 1928 GGA---CCCAGGACTCAGCAGCCAAAGCTGTTGAAGGACATCGCACTATGACACCCC 1984
QY 548 -----ProValArgTyrSerSerGluValAsnHisLeuSerProArgGluVal 564
Db 1985 CGACATCCGAAGGTTCCAGCCAGTTCCAGCAGTACGTGCGGGTATACCCAGAGAGGTG 2044
QY 565 ThrThrValLeuGlnAlaAspSerAla-GluTyr-----AlaG 577
Db 2045 GTCCACGCGGCGATCGGATGAGGCTGGAGTCTGCTGGCTGTGACTGGACAGACTCAA 2104
QY 577 nProLeuValGlyIleValGlyThrLeuHisGlnArgSerThrPheLysPro----- 595
Db 2105 GCCCAGATGGAGACCGCTGGGACCCACCTGAAGAGTGAAGACACCCACCCCTATATCC 2164
QY 596 -----GluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyr 609
Db 2165 CATGGATGAGGATGCCAGCGGTGTGGGAAACTGCAGCTTTGAGGTGACAAAGATT 2224
QY 609 rAsnSerProGly----- 613
Db 2225 GCAATTCCTTCAGGATTCACCTGCAACTTTGATTTTCTGAAGAGACCTGTGTTGGAT 2284
QY 614 -----GlnGluValTyrHisAlaTyrAlaGluPro----- 623
Db 2285 GTACGACCGTGCAGAGTGGCTGACAGTACCTGGAT-CAGCAGTGCACACCTCAACGACA 2343
QY 624 -----LeuProIleThr-----GlyPr 629

Db 2344 GAACGTTTCCAGATGACAAGAACTTCTCTGAAACTACAGACGCGGCGGACGAGGGGCC 2403
QY 629 oGluTyrAlaThr-----ProIleIleMetAspMe 639
Db 2404 AGTTTGGGCGGCTCATCAGCCACCACTGCACCTGCCCGAAGCCCTGTGTG----- 2455
QY 639 tSerGlyHisProThrThrSerValGlnProSerThrSerThrPheLysAlaThrG 659
Db 2456 -CATGAGATTCCAATACCAAGCCATGGCGGCCACCGGTGGCACTCCAGGTGTTCTGG 2514
QY 659 yAsnGlnProProLeuValGlyThrTyrAsnThrLeu-LeuSerArgThrAspSerC 679
Db 2515 AGCCAG-----ACAGGAAGCAAACTCTTTGGGTCACTCCGCGAGCACAGG 2562
QY 679 ySerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaP 699
Db 2563 CGAGCGAGTGG-----AAGCATGGACGCATATTCTGCCCAGCT 2601
QY 699 rAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArg 719
Db 2602 ATGAC---ATGAGATATCAGATC-----GTATTCCAGGGAGTGTCTGGAAAGGCGAT 2652
QY 719 spGlyGluCysAspVal 724
Db 2653 CCGGAGAGATTCTTCATC 2669

RESULT 10
US-08-936-135-17
; Sequence 17, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
US-08-936-135-17

Alignment Scores:
Pred. No.: 1.37e-29 Length: 2730
Score: 411.00 Matches: 191
Percent Similarity: 36.56% Conservative: 104

Best Local Similarity:	23.67%	Mismatches:	342
Query Match:	10.57%	Indels:	170
DB:	3	Gaps:	31
US-10-060-830-3 (1-729) x US-08-936-135-17 (1-2730)			
QY	1	MetProLeuPheLeuLeu-----LeuLeuValLeuLeuLeuLeuLeuGluAspAla	18
DB	1	ATGGATATGTTTCTCTACCTCCGTTTCTTAGCCCTCTACTTTTCAAGACACCAAGTG	60
QY	19	GlyAlaGlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGlnSerGlyThrLeu	38
DB	61	AGAGGCCAACACAGACCCCGTGGAGGTGTTGAATCCAAAGATGCTGGCTATATC	120
QY	39	ThrSerIleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluIleArg	58
DB	121	ACCTCTCCGGTTACCCCGGAGCTACCCCTCCACAGACCTCCGAGTGATGTTTAC	180
QY	59	Val---LysMetGlyGluArgValArgIleLysPheGly---AspPheAspIleGluAsp	76
DB	181	GCCCCGAACCAACAGAGATGTCCTCAACTTCAACCCCTCACTTGAATCGAGAAG	240
QY	77	SerAspSerCysHisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThr	96
DB	241	CAGGAC---TGCAGATGACTTATCGAGATTCGGGATGGGACAGTGAATCCCGAGAC	297
QY	97	GluIleGlyLysTyrCysGlyLeuGlyLeuGlnMetAsn-----HisSerIleGlu	113
DB	298	CTCCTGGCAACACTGGG-----AACATCGCCCGCCACCACCATCATC	342
QY	114	SerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGly	133
DB	343	TCCTCGGGTCCATGCTACATCAAGTTCACCTCCAGCTACCGCGGCGGCGGAGGC	402
QY	134	PheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAla	153
DB	403	TTCTCTCTCGGTACGAGATCTTCAACAGAGCTCT-----GAAGATTTGTC	450
QY	154	SerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPhe	173
DB	451	AAAACTTCAAGCCCAACGGGACCATCGAATCTCTCTGGG-----TTT	495
QY	174	AlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerSerProLeuCysMetAla	193
DB	496	CTGAG-----AAGTATCCACACACACTTG---GACTGCACCTTTACCATCTCTGCC	543
QY	194	GlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSer-----	209
DB	544	AAACCCAGATGGAGATCATCTCTGACATCTCTGATCTTTCACCTGGAGCATGCCCTTTG	603
QY	210	-----ValValIleSerLysGlyIlePro---	217
DB	604	CAGTGGGAGAGGGGACTCCAAGTACGATGGCTGGACATCTGGGATGGCATTCACAT	663
QY	218	-----TyrTyrGluSerSerLeuAlaAsnValThrSerVal	230
DB	664	GTGGCCCTGATGTCGAAGTACTGTGGGACCAACACCCCTCTGAACCTCGTTTCATCG	723
QY	231	ValGlyHisThrSerThrLeuPheThr-----	240
DB	724	ACGGGGATCTCTCCCTGACCTTCACACGGACATGCGGTGGCCCAAGGATGGCTCTCT	783
QY	241	-----PheLysThrSerGlyCysTyrGlyThrLeu	250
DB	784	CGCGGTTACTACCTGGTCCACCAAGAGCCACTAGAGAATTTTCAGTGCATGTTCTCTG	843
QY	251	GlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluTrp	270
DB	844	GGATGAGTCTGCGCGGATGCTTAATGAACAGATCATGCTGCTCATCT-----	891
QY	271	ThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLysLysPro	290
DB	892	-----ACCTACTGATGGAGGTGGACCCCTCAACAAAGCGGCTCCATGGGTGAT	942
QY	291	GlyProProTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsn	310
DB	943	GACAAATGGCTGGACCCCACTGGATTCCAAACAGAGATATCTCCAGGTGACCTGCGC	1002
QY	311	LysGluLysLysIleThrGlyIleIleThrThrGly-----SerThrMetValGluHis	328
DB	1003	TTTTTAAACATGCTCAGGCCATCCGACACAGAGGAGCAGATTTCAGGGAACACAGAAT	1062
QY	329	AsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTrpThrVal	348
DB	1063	GGCTACTAGTCAATCTACAGCTGGAAGTCAGCAGCTAATGGAGAGCTGGATGGTG	1122
QY	349	TyrArgGluProGlyValGluGlnAspLysLysIlePheGlnGlnLysAspTyrHisGln	368
DB	1123	TACCGGATGGC-----AAAAACCAACAAGTATTTCAAGCCCAACAACATGCACTGAG	1176
QY	369	AspValArgAsnAsnPheLeuProPheIleAlaArgPheIleArgValAsnProThr	388
DB	1177	GTGGTTTCTGAACAAAGTCCACGCTCCACTGCTGACAGGTTTGTAGAATCCGCCCTCAG	1236
QY	389	GlnTrpGlnGlnLysIleAlaMetLysMetGluLeuGlyCysGlnPheIleProLys	408
DB	1237	ACCTGGCACTCAGTATCGCCCTCCGCTGGAGCTCTTGGCTGC-----	1281
QY	409	GlyArgProLysLeuThrGlnProProProArgAsnSerAsnAspLeuLysAsn	428
DB	1282	-----CGGGTCACAGATGCTCCCTGCTCCAACATGCTGGGATGCTCTCAGGC	1329
QY	429	ThrThrAlaProLysIleAlaLysGlyArgAlaProLysPhe-----	443
DB	1330	CTCATTGACAGCTCCAGATCTCCGCTCTTCCACCCAGAGATACCTTGGAGCCCCAGT	1389
QY	444	ThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThr	463
DB	1390	GCAGCCCGCTGTGTCAGCAGCCCTGGGCTGGTTCCTCGATCCCTCAGGCCAGCCCC	1449
QY	464	AlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAspVal-----	481
DB	1450	GGTGAGGAGTGGCTCAGGTAGATCTGGGAACACCCCAAGACAGTGAAGGTGTCATCATC	1509
QY	482	-----AlaLeuAlaAlaValLeuValProValLeuValMetVal	494
DB	1510	CAGGAGCCCGCGGAGAGACATATCATCTGCTGGAAGCCAGACGATTTGTGGCGAAG	1569
QY	495	LeuThrThrLeuIleLeuValCysAlaTrpHisTrp-----Arg	509
DB	1570	TTCAAGTCTCTACACCTTAAACGCAAGGACTGGGAATACATTTCAGGACCCAGGACC	1629
QY	510	AsnArgLysLysLysThrGluGlyThr-----TyrAspLeuProTyrTrpAspArgAla	527
DB	1630	CAGCAGCCAAAGCTGTTGGAAGGGAACATGCATATGACACCCCT-----	1674
QY	528	GlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThr	547
DB	1675	-----GACATCCGAAGGTTGACCCCATTCGCGCAGCATGTGCGGGGTATAC	1722
QY	548	ProValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrVal	567
DB	1723	CCGAGAGGTGGTGGCGGGGATGGGATGGG-----CTGGAGGTG	1767
QY	568	LeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyIleValGlyThrLeu	587
DB	1768	CTGGGCTGTGACTGGACAGAC---TCCAAAGCCCAAGGCTGGGACCCCATGTG	1824
QY	588	HisGlnArgSerThrPheLysPro-----GluGluGlyLysGluAlaGly---	602
DB	1825	AAGAGCGAAGAGACACACCCCTTACCCACCGCAAGAGGAGCCACAGATGTGGGAG	1884
QY	603	-----TyrAlaAspLeuAspProTyrAsnSerProGly-GlnGluValTyrHisAl	619
DB	1885	AACTGCACCTTTGAGGATGACAAGATTTGCAGCTCCCTTCGGGATTCATTCGAACTTC	1944

QY 619 aTyraLagluProLeuProIleThrGlyProGluTyraLalThrProIleIleMetAspWe 639
 Db 1945 GATTTCCTCGAGGAGCCCTGTGTTGGA-----TGATGACCAT 1983
 QY 639 t-----SerGlyHisProThrSerValGlyGlnProSerThrSerThrPhely 656
 Db 1984 GCAAGTGGCTCCGGACACCTGGGCGCAGAGCTCCAGCCCAACAGCAGCGGAGTTTCCA 2043
 QY 656 salatrGlyAsnGlnProProLeuValGlyThrTyraAsnThrLeuLeuSerArgTh 676
 Db 2044 GATGACAGA-----2053
 QY 676 rAspSerCysSer-----SerAlaGlnAlaGlnTyraAspThrProLysAlaGly-L 693
 Db 2054 -ATTTCCTGGCGTGCAGAGTGACAGCCAGAGAGAGGGCCAGTATGCCCGGCTCATCAGC 2112
 QY 693 ysProGlyLeuProAlaProAspGluLeuValTyraGlnValProGlnSerThrGlnGluV 713
 Db 2113 CCCCTGTCCACCTGCTCCGCGAAGCCCGGTGTGATGGAGTTCCA-----G 2157
 QY 713 alSerGlyAlaGlyArg 718
 Db 2158 TACCAGGCCACGGGGCG 2174

RESULT 11

US-08-936-135-19
 ; Sequence 19, Application US/08936135
 ; Patent No. 6054293
 ; GENERAL INFORMATION:
 ; APPLICANT: Tessier-Lavigne, Marc
 ; APPLICANT: He, Zhigang
 ; APPLICANT: Chen, Hang
 ; TITLE OF INVENTION: Semaphorin Receptors
 ; NUMBER OF SEQUENCES: 26
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
 ; STREET: 75 DENISE DRIVE
 ; CITY: HILLSBOROUGH
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 94010
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/936,135
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: OSMAN, RICHARD A
 ; REGISTRATION NUMBER: 36,627
 ; REFERENCE/DOCKET NUMBER: UC97-288-2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (650) 343-4341
 ; TELEFAX: (650) 343-4342
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2781 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-08-936-135-19

Alignment Scores:

Pred. No.: 1,41e-29 Length: 2781
 Score: 411.00 Matches: 191
 Percent Similarity: 36.56% Conservative: 104
 Best Local Similarity: 23.67% Mismatches: 342
 Query Match: 10.57% Indels: 170
 DB: 3 Gaps: 31

US-10-060-830-3 (1-729) x US-08-936-135-19 (1-2781)
 QY 1 MetProLeuPheLeuLeu-----LeuLeuLeuValLeuLeuLeuLeuLeuLeuAspAla 18
 Db 1 ATGGATATGTTCTCCTCCTACCTCGGGTTTCTTAGCCCTCTACTTTTCAAGACACCAAGTG 60
 QY 19 GlyAlaGlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGlnSerGlyThrLeu 38
 Db 61 AGAGGCCAACCAAGACCCCGTGGGAGTGTCTTGAATTCCAAAAGATCGGCTATATC 120
 QY 39 ThrSerIleAsnTyraProGlnThrTyraProAsnSerThrValCysGluThrGluLeuArg 58
 Db 121 ACCTCTCCCGGTATACCCCGAGGACTACCCCTCCACCACCAAGTCCGAGTGGATGTTTAC 180
 QY 59 Val---LysMetGlyGluArgValArgIleLysPheGly---AspPheAspIleGluAsp 76
 Db 181 GCCCGCGAACCACCAAGAGATGTCCTCAACTTCAACCCCTCACTTTGAATTCGAGAAG 240
 QY 77 SerAspSerCysHisPheAsnTyraGlnArgIleTyraAsnGlyIleGlyValSerArgThr 96
 Db 241 CACGAC---TGCAGTATGACTTTATCGAGATTCGGGATCGGACAGTGAATCCGCGAGAC 297
 QY 97 GluIleGlyLysTyraCysGlyLeuGlyLeuGlnMetAsn-----HisSerIleGlu 113
 Db 298 CTCCTGGCAACACTGTGGG-----AACATGCCCGCCGCCACCATCATC 342
 QY 114 SerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGly 133
 Db 343 TCCTCGGCTCCATGCTCTACATCAAGTTCACCTCCGACTACGCCCGGAGGGGCGAGC 402
 QY 134 PheLeuAlaSerTyraSerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAla 153
 Db 403 TTCTCTCTCGCTACGAGATCTTCAAGACAGGCTCT-----GAAGATGCTCTCA 450
 QY 154 SerAsnPheLeuGluProGluPheSerLysTyraCysProAlaGlyCysLeuLeuProPhe 173
 Db 451 AAAAATTTCACAAAGCCCAACGGGACCATCGAATCTCCTGGG-----TTT 495
 QY 174 AlaGluIleSerGlyThrIleProHisGlyTyraArgAspSerSerProLeuCysMetAla 193
 Db 496 CDTGAG-----AAGTATCCACACAAGTTG---GACTGCACCTTTACCATCTGGCC 543
 QY 194 GlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSer-----209
 Db 544 AAACCCAAAGATGGAGATCATCTCGAGTTCTCTGATCTTTGACCTGGGAGCATGACCCCTTG 603
 QY 210 -----ValValIleSerLysGlyIlePro---217
 Db 604 CAGGTGGGAGGGGGAGTGCAGTACGATTCGCTGGACATCTGGATGGCATTCACAT 663
 QY 218 -----TyrTyraGluSerSerLeuAlaAsnValThrSerVal 230
 Db 664 GTTGCCCGCTGATTGGCAAGTACTGTGGGACCAAAACACCCCTCTCAACTTCGTTTCATCG 723
 QY 231 ValGlyHisLeuSerThrSerLeuPheThr-----240
 Db 724 ACGGGATCTCTCCTCGACCTTTTACACGGACATGGCGGTGGCCAAAGATGGCTTCTCT 783
 QY 241 -----PheLysThrSerGlyCysTyraGlyThrLeu 250
 Db 784 CGCGTTACTACCTGGTCCACCAAGAGCCACTAGAGAACTTTTCAGTGAATGTTCTCTCG 843
 QY 251 GlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluTrp 270
 Db 844 GGCATGGAGTCTGGCCGGATTGCTTAATGACAGATCAGTGCCTCATCT-----891
 QY 271 ThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLysPro 290
 Db 892 -----ACCTACTCTGATGGAGGTGGACCCCTCAACAAAGCCGGCTCCATGGTGTAT 942
 QY 291 GlyProProTrpAlaAlaPheAlaThrAspGluTyraGlnTrpLeuGlnIleAspLeuAsn 310
 Db -----

Db 943 GACAATGGCTGGACCCCAACTGGATTCACAAAGAGATATCTCCAGGTGGACCTCCGC 1002
QY 311 LysGluLysLysIleThrGlyIleIleThrGly-----SerThrMetValGluHis 328
Db 1003 TTTTAAACCATGCTCAGCGCCATCGCAACACAGGACGATTTCCAGGGAACACAGAAT 1062
QY 329 AsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspAspGlyGlnLysTyrThrVal 348
Db 1063 GGCTACTACGTCAATCTCAAGCTGGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1122
QY 349 TyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyrHisGln 368
Db 1123 TACCGCATGCG-----AAAACCAACAGATATTTCCACCAACAGATGCAACTGAG 1176
QY 369 AspValArgAsnAsnPheLeuProIleAlaArgPheIleArgValAsnProThr 388
Db 1177 GTGGTTCTGAACAAGCTCCACGCTCCACTGCTGACAAGTTTGTAGAAATCCGCGCTCAG 1236
QY 389 GlnTyrGlnGlnLysIleAlaMetLysMetGluLeuGlyCysGlnPheIleProLys 408
Db 1237 ACCTGGCACTCAGTATCGCGCTCCGCGCTCGAGCTCTTCGCGTGC----- 1281
QY 409 GlyArgProProLysLeuThrGlnProProProProArgAsnSerAsnAspLeuLysAsn 428
Db 1282 -----CGGGTCACAGATGCTCCTGCTCCAAATGCTGGGGATGCTCTCAGGC 1329
QY 429 ThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPhe----- 443
Db 1330 CTCATTGCACTCCAGATCTCGCGCTCTCCACCCAGGAATACCTCTGGAGCCCGCAGT 1389
QY 444 ThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThr 463
Db 1390 GCAGCCGCGCTGTCAGCAGCGCTCGCGCTGCTCGAATCCCTCAGGCCCGCAGCC 1449
QY 464 AlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAspVal----- 481
Db 1450 GGTGAGGAGTGGCTTCAGGTAGATCTGGGAACACCCCAAGACAGTGAAGGTGTATCATC 1509
QY 482 -----AlaLeuAlaValLeuValProValLeuValMetVal 494
Db 1510 CAGGAGCCCGCGGAGGAGACAGTATCACTGCTGTGAAGCCAGAGCATTTGTGCGGAAG 1569
QY 495 LeuThrThrLeuIleLeuLeuValCysAlaTyrHisTyr-----Arg 509
Db 1570 TTCAAGCTCTCTACACCTAAACGCGAAGGACTGGAATACATCAGGACCCCGAGGACC 1629
QY 510 AsnArgLysLysLysThrGluGlyThr-----TyrAspLeuProTyrTyrAspArgAla 527
Db 1630 CAGCAGCAAGCTGTTCGAAGGAACATGCTGATGACACCCCT----- 1674
QY 528 GlyTyrTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThr 547
Db 1675 -----GACATCCGAAGTTTGACCCCATTCGCGCACAGTATGTGCGGTATAC 1722
QY 548 ProValArgTyrSerSerGluValAsnHisLeuSerProArgGluValThrThrVal 567
Db 1723 CCGGAGAGGTGGTCCGCGCGGGATGGGATGGG-----CTGGAGGTG 1767
QY 568 LeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeu 587
Db 1768 CTGGGTGTGCTGACGACAGAC-----TCCAAGCCACGAGTGAAGAACCTGGGACCCACTGTG 1824
QY 588 HisGlnArgSerThrPheLysPro-----GluGluGlyLysGluAlaGly--- 602
Db 1825 AAGACGAGAGACACACCCCTTACCCACCCAGGAGGCGCCACAGAGTGTGGGAG 1884
QY 603 -----TyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAl 619
Db 1885 AACTGCAGCTTTGAGGATGACAAAGATTGACGCTCCCTTCGGGATTCAATTGCAACTTC 1944
QY 619 aTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetAspMe 639
Db 1945 GATTTCCTCGAGGACCCCTGTGGTGA-----TGATGACCAT 1983

QY 639 t-----SerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLys 656
Db 1984 GCCAAGTGGCTCGCGACCACTGGCCAGCAGCTCCAGGCCAACACGCGGATTTCCA 2043
QY 656 sAlaThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgTh 676
Db 2044 GATGACAGGA----- 2053
QY 676 rAspSerCysSer-----SerAlaGlnAlaGlnTyrAspThrProLysAlaGly-L 693
Db 2054 -ATTCTTGGCTGGTCAGAGTGCAGCCAGAGAGAGGCCAGTATGCCCGGCTCATCAGC 2112
QY 693 ysProGlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluV 713
Db 2113 CCCTCTGTCACCTGCCCGAAGCCGCTGTCATGGATGATTCGA-----G 2157
QY 713 alSerGlyAlaGlyArg 718
Db 2158 TACAGCCACGGGCGG 2174
RESULT 12
US-08-936-135-1
; Sequence 1, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2772 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-936-135-1
Alignment Scores:
Pred. No.: 1,44e-28 Length: 2772
Score: 400.50 Matches: 130
Percent Similarity: 42.33% Conservative: 66
Best Local Similarity: 28.08% Mismatches: 178
Query Match: 10.30% Indels: 89
DB: 3 Gaps: 18
US-10-060-830-3 (1-729) x US-08-936-135-1 (1-2772)

QY	319	IleThrThrGly-----SerThrMetValGluHIsAsnTyrTyrrValserAlaTyrArg	336
DB	1365	GGAGACACAGGGTGCCATTTCCAAGGAACCRAAGAAGAAATAATTATGTCTCAAGACTTCACAGA	1424
QY	337	IleLeuTyrSerAspAspGlyGlnLysTrpThrValTyrArgGluProGlyValcIuGln	356
DB	1425	GTAGACATCAGTCCAACCGAGGAGGACTGCCTCCCTGAAGAGGGA-----AATAAA	1478
QY	357	AspLysIlePheGlnGlnLysAsnLysAspTyrHISGlnAspValArgAsnPhelEuPro	376
DB	1479	GCCATTATCTTTCAGGGAACAACCAACCCACAGATGTTGCTTAGGAGATTTTCTCCAAA	1538
QY	377	ProileIlealaargPheIleargValasnProThrGlnTrpGlnGlnLysIleAlamEt	396
DB	1539	CCACTGATAACTGCATTTGTCGGAATCAACCTGTATCTCTGGAAACTGGTATATCTATG	1598
QY	397	LysMetGluLeuLeucGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGln	416
DB	1599	AGATTGAAGTTTATGCTGC-----AAGATAACAGAT	1631
QY	417	ProPro 418	
DB	1632	TATCCT 1637	
 RESULT 14 US-09-116-473-3			
; Sequence 3, Application US/09116473			
; Patent No. 6428965			
; GENERAL INFORMATION:			
; APPLICANT: Kolodkin, Alex			
; APPLICANT: Ginty, David			
; TITLE OF INVENTION: SEMAPHORIN RECEPTOR			
; NUMBER OF SEQUENCES: 8			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Banner & Witcoff			
; STREET: 1001 G Street, NW			
; CITY: Washington			
; STATE: DC			
; COUNTRY: USA			
; ZIP: 20001			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Diskette			
; COMPUTER: IBM Compatible			
; OPERATING SYSTEM: DOS			
; SOFTWARE: FastSeq for Windows Version 2.0			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/09/116,473			
; FILING DATE: 17-JUL-1998			
; CLASSIFICATION:			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: 60/052,762			
; FILING DATE: 17-JUL-1998			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Kagan, Sarah A			
; REGISTRATION NUMBER: 32141			
; REFERENCE/DOCKET NUMBER: 01107.74973			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 202-508-9100			
; TELEFAX: 202-508-9299			
; TELEX:			
; INFORMATION FOR SEQ ID NO: 3:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 3471 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
us-09-116-473-3			
 Alignment Scores:			
Pred. No.:	3,69e-27	Length:	3471
Score:	387.50	Matches:	192
Percent Similarity:	34.96%	Conservative:	88

QY 151 ValThrSerValValGlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCys 170
 Db 1121 GTCACTTCCACGGTGGGATATCTATCTGCAAGTCTGTTTACATTAAAGACAAGTGGTTGC 1180
 QY 171 TyrGlyThrLeuGlyMetGlySerGlyValIleAlaAspProGlnIleThrAlaSerSer 190
 Db 1181 TATGGGACTCTGGGATGGAGTCTGGTGTGATGTCCTCCATCCAGATACAGATCGTCT 1240
 QY 191 ValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArg 210
 Db 1241 GCATGGAGTGGAGTACACCATGGGCGCAGGAGAACAGCTGGACAGCGGAGAGGCCAGG 1300
 QY 211 LeuLysLysProGlyProProTrpAlaAlaPheAlaThrAspGluTrpGlnTrpLeuGln 230
 Db 1301 CTGAGAAACCGCGGCTCCTCGGCTGCTTTGCCACTGATGAGCATCAGTGGCTCGAG 1360
 QY 231 IleAspLeuAsnLysGlyLysLysIleThrGlyIleThrThrGlySerThrMetVal 250
 Db 1361 ATAGACCTTAAACAGGAGAGAGATACAGGATATCGTAACCACTGGCTACCATCATTA 1420
 QY 251 GluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspAspGlyGlnLysTrp 270
 Db 1421 GAACACAGTTACTATGTCTGCTACAGATCTCTGTACAGTACGATGGCGAGAGATGG 1480
 QY 271 ThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyr 290
 Db 1481 ACTGTGTACAGAGAACCTGGTGTGGACGAGCAAGATATTTCAAGGAAACAAAGATTAT 1540
 QY 291 HisGlnAspValArgAsnAsnPheLeuProProIleAlaArgPheIleArgValAsn 310
 Db 1541 CACAAGGATGTTCTTAACCTTTTGGCCACCAATATTGACACCTTCATTAGAGTGAAC 1600
 QY 311 ProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheIle 330
 Db 1601 CCTCTCAGTGGCAGCAAGAAATGGCATGAAGTGAACGTCGCGATGTCAGCTTACT 1660
 QY 331 ProLysGlyArgProProLysLeuThrGlnProProProArgAsnSerAsnAspLeu 350
 Db 1661 CTCAAGAGTGCCTTCCAAAGCTTACT-----CCACCTCTCGAAGCGGCAATAACCTC 1714
 QY 351 LysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPheThrGlnPro 370
 Db 1715 AGAATACTACAGCTCGTCCCAAACTAGGTAAGTGTGTCGCCCTTAATTTACTCAAGTG 1774
 QY 371 LeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerPro 390
 Db 1775 CTCCAACCTCGAAGTAGGAATGAATCTCTGTGCGAGCGCGGAGACAACTACCACTCTCT 1834
 QY 391 AspileArgAsnThrThrValThrProAsnValThrLysAspValAlaLeuAlaVal 410
 Db 1835 GATATAAACAACAGCTGTAACTCCCAAGTGTACCAAGATGTGCGCACTGGCTGCCGTT 1894
 QY 411 LeuValProValLeuValMetValLeuThrThrLeuIleLeuValLysAlaTrp 430
 Db 1895 CTGTCCTCTGTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1954
 QY 431 HisTrpArgAsnArgLysLysLysThrGluGlyThrThrAspLeuProTyrTrpAspArg 450
 Db 1955 CACTGGAGAACAGGAGAGAAACTGAAGCGCTTATGATTTACCCCACTGGGATCGG 2014
 QY 451 AlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGlu 470
 Db 2015 GCAGTTGTTGAAGGAATGAAGCACTTCTCCCTGCAAGTGTGCTGGTGGACCCAGGAG 2074
 QY 471 ThrProValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrThr 490
 Db 2075 ACGCAGTCCGCTACAGCACTAGTGAAGTCACTGAGTCCAGTCCAGGAGTCAACACA 2134
 QY 491 ValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThr 510
 Db 2135 GTGCTGCAGGCGGACTCTGCAGATATGCACAGCCCTCGTGGGAGGAATTTGTCGACA 2194
 QY 511 LeuHisGlnArgSerThrPheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeu 530

Db 2195 CTCCATCAGAGATCCACCTTTAAACCTGAGGAAGGAAGGAGGCTATGCAGACCTC 2254
 QY 531 AspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIle 550
 Db 2255 GATCCTTACAACCTCTCAATGCAGGAAGTGTACACGCTTATGCTGAACCACTGCCCGTA 2314
 QY 551 ThrGlyProGluTyrAlaThrProIleIleMetAspMetSerGlyHisProThrThrSer 570
 Db 2315 ACGGGGCTGAGTACGCAACCCGATCGTCATGCATGACATGTCAGGGGACCCACAGCTCA 2374
 QY 571 ValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProProLeuVal 590
 Db 2375 GTTGGTTCCTCCATCCACCTTCAAACTGCAGGGGACCCAGCTCAGCTTTAGTG 2434
 QY 591 GlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyr 610
 Db 2435 GGAACCTTACAACACTCTCTCTCCAGGACTGACAGCTGTTCCTCAGGCGCAGCTCAGTAT 2494
 QY 611 AspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTyrGlnVal 630
 Db 2495 GACACCCCAAAAGGTGGGAAG---TCAGCTGCTACCCAGGAGGAGTGGTATACCGAGTG 2551
 QY 631 ProGlnSerThrGlnGluValSerGlyValArgAspGlyGluCysAspValPheLys 650
 Db 2552 CCCAGAGACCCAGGAGCTATCAGGAGCAGGAGGATGAGAAGTTTGATGCTTTTAAA 2611
 QY 651 GluIleLeu 653
 Db 2612 GAAATCCTT 2620

RESULT 6

AK056350 2020 bp mRNA linear PRI 01-AUG-2002
 LOCUS Homo sapiens cDNA FLJ31788 fls, clone NT2R12008598, weakly similar
 DEFINITION to MILK FAT GLOBULE-EGF FACTOR 8 PRECURSOR.
 ACCESSION AK056350
 VERSION AK056350.1 GI:16551729
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
 clone_11b:NT2R12 clone:NT2R12008598.

ORGANISM

Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,
 Ono, Y., Hotuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,
 Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,
 Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,
 Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y.,
 Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y.,
 Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A.,
 Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
 Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 Unpublished

JOURNAL

2 (bases 1 to 2020)
 Isogai, T., Otsuki, T. and Sugiyama, T.
 Direct Submission

REFERENCE

Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction: Helix Research Institute (HRI) (supported by Japan
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
 HRI, and Biotechnology Center, National Institute of Technology and
 Evaluation; clone selection for full insert sequencing: RAB and
 HRI.

FEATURES

Location/Qualifiers
 1..2020
 /organism="Homo sapiens"

Best Local Similarity: 42.59% Mismatches: 61
Query Match: 9.58% Indels: 6
DB: 2 Gaps: 3

US-10-060-830-1113 (1-1962) x US-08-162-402B-12 (1-159)

QY 505 GCATCTTATGGAACACTGGGATGAGTCTGGTGTGATCGCGGATCCTCAATAACAGCA 564
Db 1 GlyCysAlaAsnProLeuGlyLeuLysAsnSerIleProAspLysGlnIleThrAla 20

QY 565 TCATCTGTGCTGGAG---TGGACTGACACACAGGCAAGAGAACAGTGGAAACCCAAA 621
Db 21 SerSerSerTyrLysThrTrpGlyLeuHisLeu-----PheSerTrpAsnProSer 37

QY 622 AAAGCCAGGCTGAAAAACCTGGA-----CGCCCTGGGCTGCTTTGCCACTGATGAA 675
Db 38 TyrAlaArgLeuAspLysGlnGlyAsnPheAsnAlaTrpValAlaGlySerTyrGlyAsn 57

QY 676 TACCAGTGGCTTACAATAGATTGAATAGGAAAGAAATAACAGGCAATATACCACT 735
Db 58 AspGlnTrpLeuGlnValAspLeuGlySerSerLysGluValThrGlyIleThrGln 77

QY 736 GGATCCACCATGGTGAGACAAATTACTATGTGTCTGTCTACAGAAATCCTGTACAGTGT 795
Db 78 GlyAlaArgAsnPheGlySerValGlnPheValAlaSerTyrLysValAlaTyrSerAsn 97

QY 796 GATGGCAGAAATGGACTGTGTACAGAGAGCTGGTGGAGCAAGATAAGATATTTCAA 855
Db 98 AspSerAlaAsnTrpThrGluTyrGlnAspProArgThrGlySerSerLysIlePhePro 117

QY 856 GGAACAAGATATATCACCAGGATGGCGTAATACTTTTCCCACTTATTTCACGT 915
Db 118 GlyAsnTrpAspAsnHisSerHisLysLysAsnLeuPheGluThrProIleLeuAlaArg 137

QY 916 TTTATTAGAGTGAATCCTACCAATGGCAGCAGCAAAATGGCAATGGAGCTGCTC 975
Db 138 TyrValArgIleLeuProValAlaTrpHisAsnArgIleAlaLeuArgLeuGluLeuLeu 157

QY 976 GGATGT 981
Db 158 GlyCys 159

Search completed: January 21, 2003, 09:51:29
Job time : 21.1166 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 21, 2003, 09:58:26 ; Search time 4536.98 Seconds
(without alignments)
4188.719 Million cell updates/sec

Title: US-10-060-830-1114

Perfect score: 3470

Sequence: 1 MPFLFLLLVLLVLLLEDDAGA.....TQVSGAGRGCDVFKEL 653

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US10060830/runat_16012003_091559_29008/app_query.fasta_1.1742
-DB=GenEmbl -QMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830 -CGN 1.1.6337 -runat_16012003_091559_29008 -NCPU=6 -ICPU=3
-NO_XLPY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIME=10 -YGAPEXT=0.5 -DELOP=6 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_in.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pi.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rod.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hcg_hum.*
40: em_hcg_mus.*
41: em_hcg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	3422	98.6	2939	9	AF387547 Homo sapi
2	3422	98.6	6093	9	AB073146 Homo sapi
3	3405	98.1	3858	9	BC029658 Homo sapi
4	2984.5	86.0	2339	10	AF387549 Rattus no
5	2950.5	85.0	2642	10	AF387548 Mus muscu
6	2226	64.1	2020	9	AK056350 Homo sapi
7	2126.5	61.3	1907	6	AK405795 Sequence
8	1712.5	49.4	1388	9	D29810 Human mRNA
9	922.5	26.6	192247	2	AC013497 Homo sapi
10	922.5	26.6	193623	9	AC106728 Homo sapi
11	870	25.1	2547	9	AC095973 Homo sapi
12	852.5	24.6	203102	2	AC126975 Rattus no
13	825	23.8	146515	2	AC025661 Homo sapi
14	756	21.8	1464	6	AX118816 Sequence
15	756	21.8	1620	6	AX118820 Sequence
16	756	21.8	1761	6	AX118818 Sequence
17	756	21.8	1768	6	AX118822 Sequence
18	429	12.4	192247	2	AC013497 Homo sapi
19	414	11.9	2846	10	BC026771 Mus muscu
20	367	10.6	6585	10	MMU52925 U52925 Mus musculu
21	362	10.4	7062	4	AF191308 Sus scro
22	361	10.4	6895	4	BOVFACV2A M81441 Bos taurus
23	361	10.4	6910	4	BOVFACV M14335 Human coagu
24	357.5	10.3	6893	9	HUMFAV S80643 Bos taurus
25	356.5	10.3	1895	4	S80643 X91895 B.taurus mR
26	356.5	10.3	1965	4	BTPAS67PT U70312 Homo sapien
27	355	10.2	1719	9	HSU70312 BC030828 Homo sapi
28	355	10.2	2728	9	BC030828 Sequence
29	355	10.2	4681	6	AX473349 Sequence
30	353.5	10.2	6909	6	A46255 Sequence 1
31	353.5	10.2	6909	6	A63218 Sequence 1
32	353.5	10.2	6909	6	AX071379 Sequence
33	353.5	10.2	6909	6	AX146885 Sequence
34	353.5	10.2	6909	6	AX409628 Sequence
35	353.5	10.2	6909	6	E29754 Method for
36	353.5	10.2	6909	9	HUMFAV M16967 Human coagu
37	352	10.1	2303	10	AF031524 Y11719 Bos taurus
38	348	10.0	1227	4	BTBP47PRO AX410374
39	346	10.0	1270	6	AX410374 S56151 HMFG-milk f
40	346	10.0	1270	6	S56151 HMFG-milk f
41	346	10.0	1384	6	AR081092 Sequence
42	346	10.0	1384	6	I14823 Sequence 1
43	346	10.0	1934	6	AR081093 Sequence
44	346	10.0	1934	9	HSU58516 Human breas
45	346	10.0	1987	9	AK095908 Homo sapi

ALIGNMENTS

RESULT 1

Db 1290 CGTTTATTAGATGAATCTACCAATGGCAGCAGAAAAATGCCATGAAATGGAGCTG 1349
QY 325 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProProPro 344
|||||
Db 1350 CTCGGATGTCAGTTATTCTTAAGGTCCTCCAAACCTTACTCAACCTCCACCTCCT 1409
QY 345 ArgAsnSerAsnAspLeuLysAsnThrAlaProProLysIleAlaLysGlyArgAla 364
|||||
Db 1410 CGGAACAGCAATGACCTCAAAACACTACAGCCCTCCAAAATAGCCAAAGTCGTGCC 1469
QY 365 ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 384
|||||
Db 1470 CCAAAATTACGCAACCACTACACCTCGCAGTAGCAATGATTTCTGCACACAGAA 1529
QY 385 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 404
|||||
Db 1530 CAAACAACCTGCCAGTCTGATATCAGAAATACTACCGTAACCTCAAAATGTAACCAAGAT 1589
QY 405 ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 424
|||||
Db 1590 GTAGCGTGGCTGCAGTTCTGTCCCTGTGCTCATGCTGCTCACTACTCTCATTTCTC 1649
QY 425 IleLeuValCysAlaTyrPheIleThrArgAsnArgLysLysThrGluGlyThrTyrAsp 444
|||||
Db 1650 ATATTAGTGTGCTTGGCACTGGAGAACAGAAAGAAAAAACTGAAGCACTATGAC 1709
QY 445 LeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLys 464
|||||
Db 1710 TTACTCTACTGGGACCGGGAGGTTGTGGAAAGAAATGAAGCAAGTTCTTCTCGCAAAA 1769
QY 465 AlaValAspHisGluGluThrProValArgTyrSerSerSerGluValAsnHisLeuSer 484
|||||
Db 1770 GCACTGGACCATGAGGAACCCCACTTCGCTATAGCAGCAGCAAGTAAATCACTGAGT 1829
QY 485 ProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuVal 504
|||||
Db 1830 CCAAGAGAAGTCACACAGTCTGAGCTGACTCTGCAGATGATGCTCAGCCACTGGTA 1889
QY 505 GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGlu 524
|||||
Db 1890 GGAGGAATGTTGTGTACATCTCATCAAGATCTACCTTTAAACCAGAGAAGGAAAGAA 1949
QY 525 AlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyr 544
|||||
Db 1950 GCAGGCTATGCAGACCTAGATCTTACAACTCACAGGCGAGGAAGTTTATCATGCTAT 2009
QY 545 AlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleMetAspMetSer 564
|||||
Db 2010 GCTGAACACCTCCCAATACGGGGCTGAGTATGCAACCCCAATCATCATGGACATGCA 2069
QY 565 GlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsn 584
|||||
Db 2070 GGGCACCCCAACACTTCAGTTGGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAAC 2129
QY 585 GlnProProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSer 604
|||||
Db 2130 CAACCTCCCACTAGTGGGAATACAAATACACTTCTCTCAGGACTGACAGCTGCTCC 2189
QY 605 SerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAsp 624
|||||
Db 2190 TCAGCCCAAGGCCCACTAGTATACCCGAAACCTGGGAAGCCAGGCTACCTGCCCCAGAC 2249
QY 625 GluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyValArgAspGly 644
|||||
Db 2250 GAATGGTGTACAGGTGCCACAGACACAGAAAGTATCAGGAGCAGGAAGGATGGG 2309
QY 645 GluCysAspValPheLysGluIleLeu 653
|||||
Db 2310 GAATGTGATGTTTTTAAAGAAATCCTT 2336
RESULT 2
AB073146
LOCUS

DEFINITION Homo sapiens mRNA for CLCP1, complete cds.
ACCESSION AB073146
VERSION AB073146.1 GI:20330503
KEYWORDS
SOURCE Homo sapiens lung cell_line:normal lung tissue cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 Koshikawa,K., Osada,H., Kozaki,K., Konishi,H., Masuda,A.,
Tatsumatsu,Y., Mitsudomi,T., Nakao,A. and Takahashi,T.
Significant up-regulation of a novel gene, CLCP1, in a highly
metastatic lung cancer subline as well as in lung cancers in vivo
Oncogene (2002) In press
2 (bases 1 to 6093)
Koshikawa,K., Konishi,H., Osada,H. and Takahashi,T.
Direct Submission
Submitted (16-OCT-2001) Takashi Takahashi, Aichi Cancer Center
Research Institute, Division of Molecular Oncology; Chikusa-Ku,
Kanokoden 1-1, Nagoya, Aichi 464-8681, Japan
(E-mail:takahachi-cc.jp, Tel:81-52-764-2993, Fax:81-52-764-2993)
FEATURES
Location/Qualifiers
1..6093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3q12.1-12.2"
/cell_line="normal lung tissue"
/tissue_type="lung"
1..6093
/gene="CLCP1"
/gene="CLCP1"
364..2691
/function="associated with metastatic phenotype"
/note="CUB, LCCU, coagulation factor V/VIII-homology
domains protein 1"
/codon_start=-1
/evidence=experimental
/product="CLCP1"
/protein_id="BAB91138.1"
/db_xref="GI:20330504"
/translation="MASRAVVRARRCPCQPOVRAAXAAPAAALPLSRSLPPCSNSS
FSNPLFLLLVLLLELDAGAGCGCGHTVLGPESGLTSINYQTYPSNVCEWE
IRVKGVRIRKFGDFDIEDSDCHENYLRIYNGIGVSRTEIGYCGLGQNMHSIES
KNEITLLFMGSHVSRGLFASVYIDKDLITCLDTNSFLPEPFSKYCPAGCLLP
FABISGTLPHGYRSDPLCMAGVHAGVSNLTGGQISVVISKIGPYTESLANNVSV
VGHLSLFTFTSGCYGLMGESGVADPQITASSVLEWTDHTGOENSKPKARLK
KGPFWAFAATDEYQWLOIDLAKKKITGLITTCSTMVHNHYVSARILYSDGOKW
TVTFEGVQDKRIFQGNKYHQDVRNNFLPPIIARFIRVNPQWQKRIAMNELLGCG
FIPKGRPKLTQPPRNSNDLKNTPPIAKGRAPKFTQPLQPRSSNEFFPAQTSTQ
TASPDIRNTVTNPVKDALAAVLVPLVMVLTLLILVCAHNRNRKKKTEGYD
LPWDRAGWKGKQFLPAKAVDHEETPVYSSSEVNHLSPREVTTLAQADSAEYAP
LVGSIYGLTHORSTFKPEGKEAGYADLPYNPSGOEYVHAYAEPLITPEYATPII
MLSGHPTTSVSGQPSSTFKATGNQPPPLVGYTYNTLLSRDSCSSAAQYDITPKAGP
GLPAPDELVTQVPQTSQEVSGAGRGECDFVKEIL"
577..924
/gene="CLCP1"
/note="CUB domain"
934..1191
/gene="CLCP1"
/note="LCCL domain"
1234..1710
/gene="CLCP1"
/note="F5/8 type C domain"
1942..2010
/gene="CLCP1"
/note="transmembrane domain"
4264..4269
/gene="CLCP1"
5873..5878
/gene="CLCP1"
6052..6057
/gene="CLCP1"

KTGERIRKFGDPDIEDSYCHNLKIPNGIVSRTEIGKVCYGLGLQMNQIESKGS
 ETVLPMSGIHSGRFLASYSVIDKQDLITCLDTPVSNFLEPEFSKYCPAGCLLPAE
 ISGTIPHGIRDSPLCMAGIHAGVSDVLGGQISVVISKGTPTTYESSLANNNTYISVGI
 LSTSLFTTSCYGTGLMESGVADIPQITASSVLEWTDHMGQENSKPEKARLRPG
 PPWAFATDEHQLQDLNKKKITGIVTGSTLIEHNTYVSAYRVLYLSDGQKWTY
 REPQAADKIFOGNKDYHKDVRNFPPIIARFIRVNPVQWOKIAMKVELLGCQFTL
 KRLPKLTOPPPRNNLKNNTTVHKLGRAPKFTQALOPRSRNDLPLLPACTTATPD
 VRNNTVPSPTKDVLAALVPLVLYWALITLILLYCAWHNRKRRKARCTYDLPHWD
 RAGWVKQLLPKSPVHEETPVRYNSSEVSHLSPREVTVLQADSRYTAQPLVGGI
 VSTLHORSTFKPEKEASADLDYNAPQVEYHAYAEPLVPTGPEYATPIMDMSG
 HSTASVGLSTSTFTFTAGNPPALVGTNTLLSRDSCSGOAYDTPKRGKPPAAAE
 ELVYQVQSTQOEAAGRDEKDFARETL"

BASE COUNT 612 a 606 c 594 g 527 t
 ORIGIN

Alignment Scores:

Pred. No.: 1, 97e-195 Length: 2339
 Score: 2984.50 Matches: 566
 Percent Similarity: 83.26% Conservative: 36
 Best Local Similarity: 78.28% Mismatches: 42
 Query Match: 86.01% Indels: 79
 DB: 10 Gaps: 3

US-10-060-830-1114 (1-653) x AF387549 (1-2339)

Qy 7 LeuLeuValLeuLeuLeuLeuGluAspAlaGlnGln----- 22
 Db 159 CTGCTCTTACTGCTCTACTGCTCCGGACGCTGGAGCCCAAGAGGTGATGGTGT 218
 Qy 22 ----- 22
 Db 219 GGACACACTGCTAGGCCCTGAGAGTGGAGACCTTATCATCCACTACCCACATACC 278
 Qy 22 ----- 22
 Db 279 TATCCTAACACTACTGTGTGTAATGGGAGATTCGAGTAAAGACGGGAGAGAAATTCGC 338
 Qy 22 ----- 22
 Db 339 ATCAAGTTCGGTGACTTTGACATTTGAGATTTCTGATTAATGTCACCTTAATACCTGAAA 398
 Qy 23 -----GlyLysTyrCysGlyLeuGlyLeu 30
 Db 399 ATCTTTAATGGAATGGAGTCAGCAGACGGAAATAGGCAAGTACTGTGGTCTGGTTTA 458
 Qy 31 GlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGly 50
 Db 459 CAATGAATCAGTCAATAGTCCAAAGGCAAGTGAATCAATCAAGTGTCTTTCATGAGTGA 518
 Qy 51 IleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeu 70
 Db 519 ATCCATGCTTCTGGTGGAGGATTTTGGCTTCTTACTCAGTTATAGATAAACAAGATTTA 578
 Qy 71 IleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCysPro 90
 Db 579 ATCACTTGTGGTACTGTATCAATTTTGGACCTGAGTTCAGTAACTACTGCCCA 638
 Qy 91 AlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAsp 110
 Db 639 GCTGGCTGCTGCTGCTTTGCTGAAATATCTGGAACGATTCCTCATGATATAGAGAT 698
 Qy 111 SerSerProLeuCysMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyCly 130
 Db 699 TCTTCACCGCTGTGTATGGTGAATCCATGACGAGAGTAGTGTCAAGTGTCTGGGTGGC 758
 Qy 131 GlnIleSerValIleSerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnAsn 150
 Db 759 CAATCAGCGTGTGATAGCAAGGACCCCATATACGAAGTCTTTGGCCCAACAT 818
 Qy 151 ValThrSerValGlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCys 170
 Db 819 GTCACCTCCAGTGGGATCTTATCTACGAGTCTGTTTACATTTAAGACAAAGTGTGTC 878

Qy 171 TyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSer 190
 Db 879 TATGGACCTCTAGGATGAGTGCAGTGTGTGATGCCGATCCCGATACAGATCATCT 938
 Qy 191 ValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArg 210
 Db 939 GTACTGGAGTGGACTCACCACATGGGCGAGGAGACAGCTGGAAACCCGAGAGCCAGG 998
 Qy 211 LeuLysLysProGlyProProTrpAlaIlePheAlaThrAspGluTyrGlnTrpLeuGln 230
 Db 999 CTGAGAAAACCGGGGGCTCCCTGGCTGCTTTTCCACTGATGAGCATAGTGGCTGCAA 1058
 Qy 231 IleAspLeuAsnLysGluLysLysIleThrGlyIleIleThrGlySerThrMetVal 250
 Db 1059 ATTGACCTTATAGGAGAGAGATAACAGGATCGTAACCACTGGATCTACCTGTATA 1118
 Qy 251 GluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTrp 270
 Db 1119 GAGCAACAATTACTATGTCTGCTACAGAGATTCTGTACAGTGACGATGGCAGAAATGG 1178
 Qy 271 ThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyr 290
 Db 1179 ACTGTGTACAGAGAGCGCTGCTGGGCTCAGGACAGATATTTCAAGGAAAACAAGATTAT 1238
 Qy 291 HisGlnAspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgValAsn 310
 Db 1239 CACAAGATGTTCTGTAATAACTTTTCCACCAATATTGCACGTTTCATTAGAGTGAAC 1298
 Qy 311 ProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheIle 330
 Db 1299 CCTGTCCAGTGGCAACAGAAAATGGCCATGAAAGTGAATTTGCTGGGATGTCAGTTCCAT 1358
 Qy 331 ProLysGlyArgProProLysLeuThrGlnProProProProArgAsnSerAsnAspLeu 350
 Db 1359 CTGAAAGGTGCGCTTCGAAAGCTTACTCAACCTCCCGACCTCGGAAACAGCAATAAAGCTC 1418
 Qy 351 LysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPheThrGlnPro 370
 Db 1419 ARAAACACTACAGTTCTATCCAAACTA-----GTCGTGCCCTTAATTTACTCAAGCA 1472
 Qy 371 LeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerPro 390
 Db 1473 CTCAACACCGAAGTAGGAATGACCTTCTGTCGCGCGCCAGACAACTGCCACTCCT 1532
 Qy 391 AspIleArgAsnThrThrValThrProAsnValThrLysAspValAlaLeuAlaAlaVal 410
 Db 1533 GATGTCAAAACACGACTGTGACTCCCAAGTGTGACCAAGATGTTGCACTGGCCGCGCTT 1592
 Qy 411 LeuValProValLeuValMetValLeuThrThrLeuIleLeuIleLeuValCysAlaTrp 430
 Db 1593 CTGTTCTCTGCTGGTCTATGCGCCCTCACCACATCATCTCATCTTAGTGTGCTGG 1652
 Qy 431 HistPrpArgAsnArgLysLysLysThrGluGlyThrTyrAspLeuProTyrTrpAspArg 450
 Db 1653 CATTTGGAGAAACAGAAAGAAAAAGCCGAGGACCTATGATTTTACCCCACTGGATCGG 1712
 Qy 451 AlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGlu 470
 Db 1713 GCAGGCTGGTGAAGGAGTGAAGCAGCTTCTCCCTGCAATCGGTGGAACACAGGAG 1772
 Qy 471 ThrProValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrThr 490
 Db 1773 ACGCAGTGCCTACAGCAACAGTGAAGTGTAGTCACTGACCGCCGAGGAGTACAGACA 1832
 Qy 491 ValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThr 510
 Db 1833 GTGTGCAAGCTGATTTGCAAGATACGACAGCCCTCGTGGGAGGAATTTGTTGGCACA 1892
 Qy 511 LeuHisGlnArgSerThrPheLysProGluGlyLysGluAlaGlyTyrAlaAspLeu 530
 Db 1893 CTCCATCAGAGATCTACCTTTAAACCTTGAAGAGGAAAAAGAGCGAGCTACGAGATCTA 1952
 Qy 531 AspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIle 550

01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 170005P21Rik protein.
 GN ESDN OR 170005P21RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai J., Okido T., Furuno M., Anon H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momaberts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR ENBL; AK006805; BAB24750.1;
 DR MGD; MGI:1920629; Esdn.
 SQ SEQUENCE 251 AA; 27428 MW; F86E0AA15EF51AE9 CRC64;

Alignment Scores:
 Pred. No.: 2,77e-89 Length: 251
 Score: 1151.50 Matches: 218
 Percent Similarity: 91.16% Conservatives: 9
 Best Local Similarity: 87.55% Mismatches: 21
 Query Match: 27.79% Indels: 1
 DB: 11 Gaps: 1

US-10-060-830-1 (1-2280) x Q9D9K5 (1-251)

Qy	1516	GTAAGCTGGCGTGCATGTTGTCCCTGTCGTGTCATGTCCTCATCTCATTCTC	1575
Dd	4	VallalaLeuAlaValLeuValProValLeuValMetAlaLeuThrThrLeuIleLeu	23
Qy	1576	ATATTAGTGTTGCTTGGCACTGGAGAACAAGAAAACCTGAAGGACCCTATGAC	1635
Dd	24	IleLeuValCysAlaTriHisTriArgAsnArgLysLysThrGluglyAlaTyrrasp	43
Qy	1636	TTACTTTACTGGACCGCGAGTTGGTGGAAAGGAATGAGCAGTTTCTTCCTGCARAA	1695
Dd	44	LeupToHisTriPaspArgAlaGlyTriPrtyrLysGlyMetLysGlnLeuLeuProAlalys	63
Qy	1696	GCAGTGGACCATGAGAAACCCAGTTCGCTATACACACGACGAGTAAATCACTGAGT	1755
Dd	64	ServAlaSPHsglgluthrProValArgYrSerThrSergluValserHisLeuSer	83
Qy	1756	CAAAGAGATCACACACAGTGTGTCAGGCTGACTTCGACAGATGTCGACCATGGTA	1815
Dd	84	AlaArgGluValThrThrValLeuGlnAlaaspSerAlaGluTyrrAlaGlnProLeuVal	103
Qy	1816	GGAGGAATGTTGGTACACTTCATCAAGATCTACCTTTAAACACAGAGAAGAAAGAA	1875
Dd	104	GlyGlylleValGlyThrLeuHisGlnArgSerThrPhelYsProGluGluGlyLysGlu	123
Qy	1876	GCAGCTATGCAGACTAGATCTTACAACTCACCAGGCGAGGAGTATATCATGCCAT	1935
Dd	124	AlaGlyTyrrAlaASPLeuASPProTyrrasnSerPrometGingluValtyrHisAlatyrr	143

QY 1936 GCTGAACCACTCCCAATTACGGGCGCTGAGTATGCAACCCCAATCATCATGACATGTCA 1995
 Db 144 AlaGluProLeuProValThrGlyProGluTyrrAlaThrProIleValMetAspMetSer 163
 QY 1996 GGCACCCCCAACCTTCAGTTGGTCAGCCCTCCACATCCACTTTTCAAGGCTACGGGGAAC 2055
 Db 164 GlyHisProThrAlaSerValGlyLeuProSerThrSerThrPhelYsThrAlaGlyThr 183
 QY 2056 CAACCTCCCCACCTAGTGGGAACCTTACAAATACACTTCTCTCCAGGACTGCACAGCTGCTCC 2115
 Db 184 GlnProHisAlaLeuValGlyThrTyrrasnThrLeuLeuSerArgThrAspSerCysSer 203
 QY 2116 TCAGCCCAGGCCCAGTATGATACCCGGAAGCTGGGAAGCAGGCTCTACCTCCCCCAGAC 2175
 Db 204 SerGlyGlnAlaGlnTyrraspThrProLysGlyGlys---SerAlaAlaThrProGlu 222
 QY 2176 GAATTGTGTACCAAGTGTCCACAGCAGACACAAAGAGTATCATGAGGACGAGGAGGATGGG 2235
 Db 223 GluLeuValTyrrGlnValProGlnSerThrGlnGluLeuSerGlyAlaGlyArgaspGlu 242
 QY 2236 GAATGTGATGTTTTTAAAGAAATCCTT 2262
 Db 243 LysPheaspAlaPheLysGluIleLeu 251
 RESULT 7
 Q9DAJ3 PRELIMINARY; PRT; 503 AA.
 ID Q9DAJ3 AC Q9DAJ3
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE 4631413Kilrik protein.
 GN 4631413Kilrik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Anon H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momaberts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR ENBL; AK006805; BAB24750.1;
 DR MGD; MGI:1920629; Esdn.
 SQ SEQUENCE 251 AA; 27428 MW; F86E0AA15EF51AE9 CRC64;

Alignment Scores:
 Pred. No.: 2,77e-89 Length: 251
 Score: 1151.50 Matches: 218
 Percent Similarity: 91.16% Conservatives: 9
 Best Local Similarity: 87.55% Mismatches: 21
 Query Match: 27.79% Indels: 1
 DB: 11 Gaps: 1

US-10-060-830-1 (1-2280) x Q9D9K5 (1-251)

Qy	1516	GTAAGCTGGCGTGCATGTTGTCCCTGTCGTGTCATGTCCTCATCTCATTCTC	1575
Dd	4	VallalaLeuAlaValLeuValProValLeuValMetAlaLeuThrThrLeuIleLeu	23
Qy	1576	ATATTAGTGTTGCTTGGCACTGGAGAACAAGAAAACCTGAAGGACCCTATGAC	1635
Dd	24	IleLeuValCysAlaTriHisTriArgAsnArgLysLysThrGluglyAlaTyrrasp	43
Qy	1636	TTACTTTACTGGACCGCGAGTTGGTGGAAAGGAATGAGCAGTTTCTTCCTGCARAA	1695
Dd	44	LeupToHisTriPaspArgAlaGlyTriPrtyrLysGlyMetLysGlnLeuLeuProAlalys	63
Qy	1696	GCAGTGGACCATGAGAAACCCAGTTCGCTATACACACGACGAGTAAATCACTGAGT	1755
Dd	64	ServAlaSPHsglgluthrProValArgYrSerThrSerg	

QY	1039	GGATCCACCATGGTGGAGCACAAATTACTATGTGCTGCTACAGAAATCCTGTACAGT	1098
Db	243	-----	243
QY	1099	GATGGCGCAAAATGGACTGTGTACAGAGCGCTGTGTGGACGAGATAAGATATTTCAA	1158
		:::	

South America

[illegible]

3 044

QY 1219 111A1TAGAGTGAATCC1ACCCCA1GGCAAGCAAGAA11GCCA1GTATTA1GGAGCC1GCTC 1219

DB 240 ----- 240

QY 1279 GGATGTCAGTTTATTCTTAAGGTCGTCTCCAAACCTTACTCAACCTCCACCTCCCTCGG 1338

Db 246 ----- 246

QY 1339 AACAGCAATGACCTCAAAAACACTACAGCCCCTCCAAAAATAGCCAAAGGTCGTGCCCCCA 1398

Db 246 ----- 246

QY 1399 AAATTACGCAACCACTACAACCTCGCAGTAGCAATGAATTTCTCTGCACAGACAGAACAA 1458

Db 246

QY 1459 ACAACTGCCAGTCTGTATATCAGAAATACTACCGTAACTCCAAATGTAAACCAAGATGTA 1518

Db 247 -----AsnIleThr-----ThrVal 251

QY 1519 GCGCTGGCTGCAGTTCTTGTCCCTGGTCATGGTCTCACTACTCTCATTTCTCATA 1578

Db 252 AlaIleproSerValIle-----pheIleAlaLeuLeuLeuThrglyMetGlyIlephe 269

OV 1579 TTAGTGTGTGCTTGGCCTGGAGAAACAGAAAGAAAAAAGGACCTATGACTTA 1638

270 AlalTleCys-----ArgIysArgIysIysIysGlyAsp----- 280

ov 1639 CCTTAC-----TGGGACCGGGCAGGTTGGTGGAAAGGAATGAAGCAGTTTCTTT 1686

298

1607 1740

[illegible]

.....

[illegible]

DO NOT WRITE IN THESE SPACES

```
QY I849 ACCIIIAACCA ----- GAAAGGA 1000
```

DB 349 TnIPheArgProMetAspTnHAspTnIGluGluIuValAlaGValAsnIuIuGluAlaSerGly 300

QY 1882 TATGCAGACC TAGATCCCTTACAACTCACCAGGGCAGGAAGTTTATCATGCCATATGCTGAA 194

Db 369 His-----TyrAspCysProHisArgProGlyArgHisGluTyrAlaLeu 383

1942 CCAC TCCCAATTACGGGGCCTGAGTATGCAACCCCAATCATC----- 1983

Db 384 ProLeuThrHisSerGluProGluTyrAlaThrProIleValGluArgHisLeuLeuArg 403

Qy 1984 -----ATGGACATGTCAGGGCACCCCAACTTCAGTTGGTCAGCCCTCCACA 2031

Db 404 AlaHisThrPheSerThrGlnSerGlyTyrArgValProGly--ProArgProThrHis 422

QY 2032 TCCACTTTCAGGCTACGGGGAACCAACCTCCCCCACTAGTGGAACTTACAATACACTT 2091

```

Db 423 LysHisSerHisSerGlyGlyPheProAlaThrGlyAlaThrGlnValGluSer 442
QY 2092 CTCCTCAGAGTACAGAGTGTCTCTCCAGCCAGCCAGTATGATACCCGAAAGCTGGG 2151
Db 443 TyrGlnArgProAlaSerProLysProValGlyGlyGlyTyrAspLysPro---AlaAla 461
QY 2152 AAGCCAGGTCTACCTGCCAGAGCAAGTGTGTACCAAGGTGCCAGAGCAGACACAAAGAA 2211
Db 462 SerSerPheLeuAspSerArgAsp-----ProAlaSerGlnSerGln 475
QY 2212 GPATCAGGAGCAGGAAGGATGGG 2235
Db 476 MetThrSerGlyGlyAspGly 483

RESULT 8
Q9D696 PRELIMINARY; PRT; 460 AA.
AC Q9D696;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE 4631413K11Rik protein.
GN 4631413K11Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=SKIN;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Buit C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; AK014521; BAB29409.1; -.
DR MGD; MGI:1913936; 4631413K11Rik.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
SQ SEQUENCE 460 AA; 50334 MW; 3CF3356F70A8B93 CRC64;

```

Alignment Scores:

Pred. No.:	2,08e-34	Length:	460
Score:	507.50	Matches:	169
Percent Similarity:	35.11%	Conservative:	81
Best Local Similarity:	23.74%	Mismatches:	161
Query Match:	12.25%	Indels:	301
DB:	11	Gaps:	21

US-10-060-830-1 (1-2280) x Q9D696 (1-460)

QY 187 CTTACATCCATAAATACCACAGAGCTATCCCAACAGCAGCTGTTGTGAATGGGAGATC 246

```

Db 1 MetThrSerLysAsnTyrProGlyThrTyrProAsnTyrThrValCysGluLysIlelle 20
QY 247 CGTGTAAAGATGGGAGAGAGTTCGCATCAAAATTTGGTGTGACTTTTGACATTGAAATCT 306
Db 21 ThrValProLysGlyLysArgLeuIleLeuArgLeuGlyAspLeuAsnIleGlu---Ser 39
QY 307 GATCTCTGTCACTTAAATTTACTTGTGAGAAATTTATAATGGAATTTGGAGTGCAGAGAACTGAA 366
Db 40 LysThrCysAlaSerAspTyrLeuLeuPheSerSer-----AlaThrAspGln 55
QY 367 ATAGCAAAATACTGTGTCTGGGGTTGCAATGAACCAATCAATTAATCAAAAGCAAT 426
Db 56 TyrGlyProTyrCysGly---SerTrpAlaValProLysGluLeuArgLeuAsnSerAsn 74
QY 427 GAATACATTCCTGCTTCATGAGTGAATCCATGTTTCTGGAGCGGATTTTGGCTCA 486
Db 75 GluValThrValLeuPheLysSerGlySerHisIleSerGlyArgGlyPheLeuLeuThr 94
QY 487 TACTCTGTTATAGATAAACAAGATCTAATTTACTTTGTGGACACTGCATCCCAATTTT 546
Db 95 TyrAlaSerSerAspHisProAspLeuIleThrCysLeuGluArgGlySerHisTyrPhe 114
QY 547 GAACCTGAGTTCAGTAAGTACTGCCAGTGTGTCTGCTTCTTCTTCTGCTGAGATATCT 506
Db 115 GluGluLysTyrSerLysPheCysProAlaGlyCysArgAspIleAlaGlyAspIleSer 134
QY 607 GGAAACAATTCCTCATGATATAGATGATTCCTCGCCATCTGTCATGCTGCTGTCATGCA 666
Db 135 GlyAsnThrLysAspGlyTyrArgAspThrSerLeuLeuLysAlaIleHisAla 154
QY 667 GCAGTAGTGTCAACACACAGTGGCGGCCCAATCACTGTTGTATAGTAAGATATTCCT 726
Db 155 GlyIleIleThrAspGluLeuGlyGlyHisIleAsnLeuLeuGlnSerLysGlyIleSer 174
QY 727 TATTGAAGAATTCCTTGGCTTAACAGTCACATCTGTGGTGGACACATTTACTACAAGT 786
Db 175 HistyGluGlyLeuLeuAlaAsnGlyValLeuSerArgHisGlySerLeuSerGluLys 194
QY 787 CTTTTTACATTTAAACACAGTGTGATTTATGGAACACTGGGATGGAGTGTGTGTGATC 846
Db 195 ArgPheLeuPheThrThr----- 200
QY 847 GCGGATCCTCAATAACAGCATCATCTGTGCTGGAGTGGACTGACCACACAGGCAAGAG 906
Db 200 ----- 200
QY 907 AACAGTTGAAACCCAAAAGCCAGGCTGAAAAAACCTGGACCGCTTGGGCTGCTTT 966
Db 200 ----- 200
QY 967 GCCACTGATGATACCACTGGTTACAAATAGATTGTAATAGGAAAGAAATAACAGGC 1026
Db 200 ----- 200
QY 1027 ATTATAACCACTGGATCCACCATCCACCATGTTGGAGACACAATTTACTGTCTGCTACAGATC 1086
Db 200 ----- 200
QY 1087 CTGTACAGTGTATGGGCGAGAAATGGAGTGTGTACAGAGAGCTGTGTGGAGCAAGAT 1146
Db 201 ----- 203
QY 1147 AAGATATTTCAAGGAAACAAGATTTATCACCAGGATGTGGTGAATTAACITTTTGGCCACCA 1206
Db 203 ----- 203
QY 1207 ATTATTGCACGTTTATTATTAGTAGTAATCCTACCCCAATGGCAGCAGAAATTTGCCATGAAA 1266
Db 203 ----- 203
QY 1267 ATGGAGCTGCTCGGATGTCAGTTTATTCCTAAAGGTGCTCTCCAAACACTTACTCAACCT 1326

```


Alignment Scores:

Pred. No.: 2,72e-26 Length: 432
Score: 412.00 Matches: 145
Percent Similarity: 33.08% Conservative: 71
Best Local Similarity: 22.21% Mismatches: 141
Query Match: 9.94% Indels: 296
DB: 11 Gaps: 19

US-10-060-830-1 (1-2280) x Q9R327 (1-432)

```
QY 364 GAAATAGGCAATACGTGCTGGTTCGCAATGAACCAATCAATGAATCAAAAGGC 423
   |||:|||||
Db 27 GluLeuGlyProFyrCysGly----SerTIPAlaValProLysGluLeuArgLeuAsnSer 45

QY 424 AATGAATACATGCTGTTTCATGAGTGAATCCATGTTTCGGAGCGGATTTTGGCC 483
   |||:|||||
Db 46 AsnGluValThrValLeuPheLysSerGlySerHisLeuSerGlyArgGlyPheLeuLeu 65

QY 484 TCATACCTCTTTATAGATAACAAGATCAATTAATCTGTTGGACACTGCATCAATTTT 543
   |||:|||||
Db 66 ThrTyrAlaSerSerAspHisProAspLeuLeuThrCysLeuGluArgGlySerHisTyr 85

QY 544 TTGGAACCTGAGTTCAGTACTGCCAGCTGGTGTGCTTCTCTTTTGTGCTGAGATA 603
   |||:|||||
Db 86 PheGluGluLysTyrSerLysPheCysProAlaGlyCysArgAspIleAlaGlyAspIle 105

QY 604 TCTGGACAAATTCCTCATGATATAGATTCCTCCGCTTGTGCTGCTGCTGCTGCTG 663
   |||:|||||
Db 106 SerGlyAsnThrLysAspGlyTyrAspThrSerLeuLeuLeuLysLysAlaIleHis 125

QY 664 GCAGGAGTAGTGTCAACACAGTGTGGCGCCAAATCAGTGTGTTAATAGTAAAGTATT 723
   |||:|||||
Db 126 AlaGlyIleIleThrAspGluLeuGlyGlyHisIleAsnLeuLeuGlnSerLysGlyIle 145

QY 724 CCTATTATCAAGATCTTTGGCTACAACGTCACATCTGCTGCTGGACACTTATCTACA 783
   |||:|||||
Db 146 SerHisTyrGluGlyLeuLeuAlaAsnGlyValLeuSerArgHisGlySerLeuSerGlu 165

QY 784 AGTCTTTTACATTTAAGACAAGTGGATGTATTGGAACACTGGGATGGAGTCTGTGTG 843
   |||:|||||
Db 166 LysArgPheLeuPheThr----- 172

QY 844 ATCGCGGATCCTCAATAACAGCATCATCTGTGCTGGAGTGGACTGACACACAGGCGAA 903
   |||:|||||
Db 172 ----- 172

QY 904 GAGACAGTTTGGAAACCCAAAGACCGAGCTGAAACAACTGGACCGCTTGGGCTGCT 963
   |||:|||||
Db 172 ----- 172

QY 964 TTGCCCACATGATGATACCAGTGGTTACAAATAGATTGAATAGAAAGAAATAACA 1023
   |||:|||||
Db 172 ----- 172

QY 1024 GGCATTATAACCACTGGATCCACCATGGTGGAGACAAATTAATGTTGCTGCCTACAGA 1083
   |||:|||||
Db 172 ----- 172

QY 1084 ATCCTGTACAGTATGATGGCAGAAATGGACTGTGTACAGAGAGCTGTGTGGAGCAA 1143
   |||:|||||
Db 173 -----ProGlyMet----- 175

QY 1144 GATAAGATATTCAAGAAACAAGATATACACAGGATGTGCGTAAATACTTTTGCCA 1203
   |||:|||||
Db 175 ----- 175

QY 1204 CCAATTATTGCAGCTTTTATTAGAGTGAATCTACCAATGGCAGCAGAAATTTGCCATG 1263
   |||:|||||
Db 175 ----- 175

QY 1264 AAAATGGAGTGTCTGGATGTCAGTTTATTCTAAAGTGTCTCTCCCAAACTTACTCAA 1323
   |||:|||||
Db 175 ----- 175
```

```
QY 1324 CCTCCACCTCCTCGGAACAGCAATGACCTCAAAACACTACAGCCCTCCAAAAATAGCC 1383
   |||:|||||
Db 175 ----- 175

QY 1384 AAAGTGTGTGCCCAAAATTTAGCAACCACTACAACTCGCAGTAGCAATGAATTTCT 1443
   |||:|||||
Db 175 ----- 175

QY 1444 GCACAGACAGAACAACTGCGCAGTCTGATATCAGAAATACTACCGTAACCTCAAT 1503
   |||:|||||
Db 176 -----Asn 176

QY 1504 GTAACAAAGATGATAGCTGGCTGCAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCT 1563
   |||:|||||
Db 177 IleThr---ThrValAlaIleProSerValIle-----PheIleAlaLeuLeuLeuThr 193

QY 1564 ACTCTCATCTCTATATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1623
   |||:|||||
Db 194 GlyMetGlyIlePheAlaIleCys-----ArgLysArgLysLysLys 207

QY 1624 GGCACCTATGACTTACTTAC-----TGGACCGGCGCAGGTTGGTGAAGGA 1671
   |||:|||||
Db 208 GlyAsn-----ProTyrValSerAlaAspAlaGlnLysThrGlyCysTrrPlysGln 224

QY 1672 ATGAACAGTCTTCTCTGCAAAAGCAGTGGACCATGAGGAACCCCA---GTTCTGCTAT 1728
   |||:|||||
Db 225 IleLysTyr-----ProPheAlaArgHis 232

QY 1729 AGCAGCAGGAA-----GTTAATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1773
   |||:|||||
Db 233 GlnSerThrGluPheThrIleSerTyrAspAsnGluLysGluMetThrGlnLysLeuAsp 252

QY 1774 GTCTCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1833
   |||:|||||
Db 253 LeuIleThrSerAspMetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrVal 272

QY 1834 CTTTCATCAAGATCTACCTTTAAACCA-----GAAGAA----- 1866
   |||:|||||
Db 273 AlaArgLysGlySerThrPheArgProMetAspThrAspThrGluValArgValAsn 292

QY 1867 GGAAGAACAGCAGCTATGACAGCTAGCTTACAACTCACCAGGCGCAGGAGTTTAT 1926
   |||:|||||
Db 293 ThrGluAlaSerGlyHis-----TyrAspCysProHisArgProGlyArg 307

QY 1927 CATGCTCTATCTCAACCACTCCCAATTCAGGGGCTGAGTATGCAACCCCAATCATC--- 1983
   |||:|||||
Db 308 HisGluTyrAlaLeuProLeuThrHisSerGluProGluTyrAlaThrProIleValGlu 327

QY 1984 -----ATGACATGTCTAGGGCACCCACCACTTCAGTT 2016
   |||:|||||
Db 328 ArgHisLeuLeuArgAlaHisThrPheSerThrGlnSerGlyTyrArgValProGly--- 346

QY 2017 GGTACGCCCTCCACATCCACTTTCAGGCTACGGGGAACCACTCCCCCTAGTGGGA 2076
   |||:|||||
Db 347 ProArgProThrHisLysHisSerHisSerGlyGlyPheProAlaThrGlyAla 366

QY 2077 ACTTACAATACATCTCTCCAGGACTGACAGCTGCTCTCAGCCAGGCGCAGTATGAT 2136
   |||:|||||
Db 367 ThrGlnIleGluSerTyrGlnArgProAlaSerProLysProValGlyGlyTyrAsp 386

QY 2137 ACCCGAAGCTGGGAAGCCAGGCTTACCTGCCCGCAGCAATTTGGTGTACAGGTGCCA 2196
   |||:|||||
Db 387 LysPro---AlaAlaSerSerPheLeuAspSerArgAsp-----Pro 399

QY 2197 CAGACACACAAGATGATCAGGACGAGGAGGAGGATGGG 2235
   |||:|||||
Db 400 AlaSerGlnSerGlnMetThrSerGlyGlyAspGly 412
```

RESULT 13
Q96I90
ID Q96I90 PRELIMINARY; PRT; 609 AA.
AC Q96I90;

DT	01-DEC-2001 (TREMBLrel. 19, Created)
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Similar to neuropilin 1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxId=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=MUSCLE;
RA	Strausberg R.;
RL	Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; BC007737; AAH07737.1; -.
DR	InterPro; IPR000859; CUB_Gdomain.
DR	InterPro; IPR000421; FA58_C.
DR	Pfam; PF00431; CUB; 2.
DR	Pfam; PF00754; F5_F8_type_C; 2.
DR	PROSITE; PS01180; CUB; 2.
DR	PROSITE; PS01285; FA58C_1; UNKNOWN_2.
DR	PROSITE; PS01286; FA58C_2; UNKNOWN_2.
SQ	SEQUENCE 609 AA; 68347 MW; 798AAFD2D568C127 CRC64;
 Alignment Scores:	
Pred. No.:	2,36e-25 Length: 609
Score:	401.50 Matches: 133
Percent Similarity:	41.41% Conservative: 67
Best Local Similarity:	27.54% Mismatches: 180
Query Match:	9.69% Indels: 103
DB:	Gaps: 19
 US-10-060-830-1 (1-2280) x Q96I90 (1-609)	
QY	31 CTCTCCGGTCCCTCCCTCCCTGCTCCCAACTCCTCCTCTCTCCATGCCTCTGTTCCTC 90
Db	1 MetcLuArggLyLeuPro----- 6
QY	91 CTGCTCTTACTTGCTGCTGCTCGTCGAGGACGCTGGAGGCCAGCAAGGTGATGGA 150
Db	7 LeuLeuCysAlaValLeuAlaLeuValLeuAlaProAlaGlyAlaPheargAsnAspLys 26
QY	151 TGTGGCACACTGCTACTAGGCCCTGAGGTGGAACCTTTACATCCATAAACCACACAG 210
Db	27 CysGlyAspThrIleylSylleGluSerProGlyTyTyLeuThrSerProGlyTyProHis 46
QY	211 ACCTATCCCACAGCACTGTTGTGAATGGGAGATCCGGTGA---AAGATGGGAGAGA 267
Db	47 SerTyHisProSerGluLysCysGluTipLeuileGlnAlaProAspProTyGlnArg 66
QY	268 GTTCGCATCAAATTGGT---GACTTTGACATGAGATCTGATCTTGTTCACITTAAT 324
Db	67 IleMetIleasnPheasnProHisPheAspLeuGlusPargAsp---CyslSyrAsp 85
QY	325 TACTTGAGAATTTATATGAATTTGGAGTCAGACAGCAACTGAATAGGCAAACTACTGTGT 384
Db	86 TyrValGluValPheaspGlyGluasngLysGlyHisPheargglyLysPheCys--- 104
QY	385 CTGGGGTGGCAATGAACCATTCGAATGAATCAAAAGCGCAATGAATACATGCTGCTTC 444
Db	105 ---GlyLysIleAlaProProProvalValSerSerGlyProPhePheIleyLysPhe 123
QY	445 ATGATGGAATCCATGTTTCTGCAGCGGATTTTTGGCTCATACCTCTGTATAGATAAA 504
Db	124 ValSerAspTyTyGluThrHisGlyAlaglyPheSerIleargtyrGluIleyLysArg 143
QY	505 -----CAAGATCAATTAATCT- 519
Db	144 GlyProGluCysSerGlnasnTyThrThrProSerGlyValIleyLysSerProGlyPhe 163
QY	519 ----- 519
Db	164 ProGluLysTyTyProAsnSerLeuGluCysThrTyTyIleValPheAlaProLysMetSer 183

[illegible]

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	417.5	10.1	925	4	US-09-116-473-2	Sequence 2, Appl	
2	414.5	10.0	901	3	US-08-936-135-22	Sequence 22, Appl	
3	414.5	10.0	906	3	US-08-936-135-24	Sequence 24, Appl	
4	414.5	10.0	909	3	US-08-936-135-8	Sequence 8, Appl	
5	412.5	10.0	909	3	US-08-936-135-10	Sequence 10, Appl	
6	412.5	10.0	914	3	US-08-936-135-12	Sequence 12, Appl	
7	412.5	10.0	926	3	US-08-936-135-14	Sequence 14, Appl	
8	412.5	10.0	931	3	US-08-936-135-16	Sequence 16, Appl	
9	398.5	9.6	909	3	US-08-936-135-18	Sequence 18, Appl	
10	398.5	9.6	926	3	US-08-936-135-20	Sequence 20, Appl	
11	398.5	9.5	923	3	US-08-936-135-6	Sequence 6, Appl	
12	398.5	9.4	922	4	US-09-116-473-4	Sequence 4, Appl	

Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSerSer 241
QY 766 GTGGGACACTTATACAAAGTCTTTTACATTTAAGACAAGT----- 807
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 808 -----GGATGTAT 816
Db 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProGluAsnPheGlnCysAsn 278
QY 817 GGAACACTGGGATGGAGTCTGGTGTGATCGGGATCTCAATAACACAGCATCTCTGTG 876
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer--- 297
QY 877 CTGGAGTGCATGACACACAGGCGGAAGAGACAGTGTGAAACCCAAAAGCCAGGCTG 936
Db 298 -----ThrPheSerAspGlyArgTyrThrProGlnGlnSerArgLeu 311
QY 937 AAAAAACCTGGACCGCTTGGCTGCTTTGGCCACTGATGAATACCATGTTTAAATA 996
Db 312 HisGlyAspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnVal 331
QY 997 GATTTCAATAAGAAAAGAAAATAACAGGCATTATAACACCTGGA-----TCCACCATG 1050
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGGACACAATTACTATCTGCTGCTACAGAACTCTGTACAGTATGATGGGAGAAA 1110
Db 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1111 TGGACTGTGTACAGAGCCTGGTGGAGCAAGATAGATATTTCAAGGAACAAGAT 1170
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsnAsp 389
QY 1171 TATCAGCAGATGTGGTAAATAACTTTTGGCCCAATAATTATTCACAGTATTATAGAGTG 1230
Db 390 AlaThrGluValAlaLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1231 AATCTACCAATGGCAGCAGAAAATTGCCATGAAGTGGAGCTGCTCGATGTGATGTT 1290
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1291 ATTCCTAAGGTGCTCTCAAACTTACTCAACCTCCACCTCCTCGGAACAGCAATGAC 1350
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCAAAACACTACAGCCCTCAAAATAGCCAAAGTGTGCCCCAAATTT----- 1404
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTyr 460
QY 1405 -----ACGCAACACTACAACTCGCAGTAGCAATTAATTTCTCGCAGACAGAA 1455
Db 461 SerProSerAlaAlaArgLeuValSerArgSerGlyTrpPheProArgAsnProGln 480
QY 1456 CAACAACCTCCAGTCTGTATATCAGAAATACTACCGTAACTCCAAATGTAACCAAGAT 1515
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----GCGCTGGCTGCAGTTCTTGTCCCTGTGCTG 1548
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1549 GTCATGCTCTACTACTCTCATCTCATATATAGTGTGCTTGGCAGCTGG----- 1599
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1600 -----AGAACACGAAGAAAACACTGAGGCACC-----TATCACTTACTTCTG 1647
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1648 GACCGGCAGGTTGGTGAAGAAATGAAGCAGTCTTCTTCCAAAGCAGTGGACCAT 1707
Db 559 -----AspIleArgArgPheAspPro----- 565

QY 1708 GAGGAAACCCAGTTCGTATATAGCAGCAGCGAGGAGTAAATACCTCAGTCCACAGAAAGTC 1767
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTyrPsrProAlaGlyIle 583
QY 1768 ACCACAGTGTGCGGCTGACTCTGCAGATAT-----GCTCAGCCACTGTGTAGGAGA 1821
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1822 ATTGTGTGTACACTTCAATCAAGATCTACCTTTAAACCA-----GAAGAAGA 1869
Db 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAGNAGCAGGCTAT-----GCAGACCTAGATCTTCAACTCA 1908
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerG1 643
QY 1909 CCAGGCGCAGGAAGTTTATCATGCTT-----ATGCTGAACCACTCCCA 1950
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTyrValTyrAspHisAlaLys 663
QY 1951 A---TTACGGGCGCTGAGTATGCAACCCCAATCATCATGATGTCAGGGCACCCCA--- 2005
Db 663 stripLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 2006 -----CAACTTCAG----- 2014
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspLysArgG1 696
QY 2015 -----TTGGTCAGCCCTCCACATCCACTTTCAAGCTACGGGGAACCAA 2058
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 2059 COTCCCGCAGTGTGGAACTTACATACACTTCTCTC----- 2096
Db 711 rgsrProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2097 -----CAGGACTGACAGCTGCTC-----CTCAGCCAGGCCAG 2130
Db 731 InValValArgGluAlaSerGlnGlnSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2131 TATGATACCCCAAGGCTGGGAGCCAGTCTACTCTGCCCCAGCAATTTGTGTACCAG 2190
Db 751 GlySerGluTrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2191 GTCCACAGCAGCACACAAAGATATCAGCAGGAGGAGGATGGGAATGTGTAT 2247
Db 770 Ile-----ValPheGluGlyValIleGlyLysArgSerGlyGluIleSerIle 786

RESULT 3

US-08-936-135-24
Sequence 24, Application US/08936135
Patent No. 6054293

GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: He, Zhigang
APPLICANT: Chen, Hang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,135

```

; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-936-135-24

Alignment Scores:
Pred. No.:      8.5e-30      Length:      906
Score:          414.50      Matches:      205
Percent Similarity: 36.59%   Conservative: 110
Best Local Similarity: 23.81% Mismatches:    336
Query Match:     10.00%     Indels:       210
DB:              3         Gaps:         37

US-10-060-830-1 (1-2280) x US-08-936-135-24 (1-906)
QY 76 ATGCCTCTGTCCTCTGCTCTTACTTGTCTGCTGCTGCTGCTC-----GAGGACGGT 129
Db 1 MetAspMetPheProLeuThrTrpValPheLeuAlaLeuTyrPheSerGlyHisGluVal 20
QY 130 GGAGCCCGCAGGAGTGATGGATGGACACACTGACACTGACTAGGCCCTGAGAGTGGACCCCT 189
Db 21 ArgSerGlnGlnAspProProCysGlyGlyArgProAsnSerLysAspAlaGlyTyrIle 40
QY 190 ACATCCATAAACTACCCACAGACCTATCCCAACAGCAGCTGTTTGTGAATGGGAGATCCGT 249
Db 41 ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluTrpIleValTyr 60
QY 250 GTA---AGATGGGAGAGAGAGTTCGCATCAAAATTTGGT---GACTTTGACATTTGAAGAT 303
Db 61 AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGluLys 80
QY 304 TCTGATTTCTGCTCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 363
Db 81 HisAsp---CysLysTyrAspPheIleGluIleLeuLeuLeuLeuLeuLeuLeuLeuLeu 99
QY 364 GAAATAGGCAATACTGCTGGTCTGGGTTGCAAAATGAAC-----CATTCAATTGAA 414
Db 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProThrIleIle 114
QY 415 TCAAAAGGCAATGAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 474
Db 115 SerSerGlySerValLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly 134
QY 475 TTTTGGCCTCATCTCTTTATAGATAACAAGATCAATTAATTAATTAATTAATTAATTAATTA 534
Db 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 535 TCCAAATTTTGGACCTGAGTTCAGTACAGTACTGCCAGCTGGTGTCTGCTCTCTTTT 594
Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly:-----Phe 165
QY 595 GCTGAGATATCGGAACAATTCCTCATGATATAGATATTCCTGCGCATTTGTCAGTGGCT 654
Db 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 655 GGTGTGCATGACGAGTAGTGTCAACACACGTTGGCGGCGCAATCATGT-----702
Db 182 LysProA-gmetGluIleIleLeuGlnPheLeuThrPheAspLeuHisAspProLeu 201
QY 703 -----GTTGTAATTAAGTAAAGGTATTCCTCC---726

```

```

Db 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis 221
QY 727 -----TATTATGAAGTTCTTTGGCTAAACAAGCTCACACTCTGTG 765
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSer 241
QY 766 GTGGGACACACTTATACAAAGTCTTTTACATTTAAGACAAGT-----807
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 808 -----GGATGTAT 816
Db 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProProGluAsnPheGlnCysAsn 278
QY 817 GGAACACTGGGATGGAGTCTGGTGTGATCGCGGATCTCAATAACAGCATCACTCTGTG 876
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSer 297
QY 877 CTGGAGTGGACTGACCACACAGGCAAGACAAGTGTGAAACCCCAAAAAGCCAGGCTG 936
Db 298 -----ThrPheSerAspGlyArgTyrThrProGlnGlnSerArgLeu 311
QY 937 AAAAAACCTGGACCCCTTGGCTGCTTTTGGCCTGATGATACACAGTGGTTACAAATA 996
Db 312 HisGlyAspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnVal 331
QY 997 GATTTGAATAAGAAAGAAATAACAGGCATTAATACCATCTGGA-----TCCACCATG 1050
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGGAGCAAAATTAATGTGCTGCTGCTACAGAACTCTGTACAGTATGATGGCGAGAAA 1110
Db 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuValSerThrAsnGlyGluAsp 371
QY 1111 TGGACTGTGTACAGAGAGCCTGTGTGGGCAAGATAAGATATTTCAAGAAACAAGAT 1170
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsp 389
QY 1171 TATCACCAGGATGGGTAAATTAATTTTGGCCACCAATTAATTTGACAGTGTATTAGATG 1230
Db 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1231 AATCTTACCAATGGCAGCAGAAAATTCATGAAATGAGCTGCTCGGATGTCAGTTT 1290
Db 410 ArgProGlnThrThrPheLysLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys 427
QY 1291 ATTCTTAAAGGTCGCTCTCCAAACTTACTCAACCTCCACCTCCTCGGAACAGCAATGAC 1350
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCAAAAACATACAGCCCTCCAAAATAGCCAAAGGTGGTGGCCCAAAATTT-----1404
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGlyLeuTrp 460
QY 1405 -----ACGCAACCACTACAACCTCGCAGTAGAATGAATTTCTCCACAGACAGAA 1455
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTyrPheProArgAsnProGln 480
QY 1456 CAACAACCTGCCAGCTCTGATATACGAATACTACCGTAACCTCAAAATGTAAACCAAGAT 1515
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----GCCTGGCTGGAGTCTTCTGCTCCCTGTGCTG 1548
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1549 GTCATGGTCTCTACTCTCTCTTCTCATTTATAGTGTGTGTGCTGCTGCTGCTGCTGCTG 1599
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1600 -----AGAAACAGAGAAAGAAAGAAAGTGAAGGCACC-----TATGACTTACTTACTGG 1647

```


QY 655 GGTGTCAGCAGGAGTAGTGTCAACACAGCTTGGGGGCCCAATACAGT----- 702
Db 182 LysProArgMetGluIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu 201
QY 703 -----GTTCTAATTAGTAAGGTATTCC----- 726
Db 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleThrPheGlyIleProHis 221
QY 727 -----TATTATGAAGTCTTGGCTAACAACGTCACATCTGTG 765
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSer 241
QY 766 GTGGACACTTACTACAACTCTTTTACATTTAAGCAAGT----- 807
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 808 -----GGATGTAT 816
Db 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProProGluAsnPheGlnCysAsn 278
QY 817 GGAACACTGGGGATGGAGTCTGTGTGATCGCGATCTCAAAATAACAGCATCATCTGTG 876
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSer 297
QY 877 CTGGAGTGGACTGACACACAGGCAAGACAGATTTGGAAACCCAAAGCCAGCGTG 936
Db 298 -----ThrPheSerAspGlyArgTrpThrProGlnGlnSerArgLeu 311
QY 937 AAAAACCTGGCCGCTTGGCTGCTTTGCCACTGATGATACCATCAGTGGTTCAATA 996
Db 312 HisGlyAspAspAspGlyTrpThrProAsnLeuAspSerAsnLysGluTyrLeuGlnVal 331
QY 997 GATTGTAATAAGGAAGAAATAACAGCATTTAACCCTGGA-----TCCACCATG 1050
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGGACACAATTAATGTGTCTGCTCAGCAATCTGTACAGTATGATGGCAGAAA 1110
Db 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1111 TGGACTGTGACAGAGCCTGTGTGGAGCAAGATAGATATTTCAGGAACAAGAT 1170
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsp 389
QY 1171 TATCACCAGATGTCGCTAATAACTTTTGGCCCAATATTATTCACAGTGTATTAGAGTG 1230
Db 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1231 AATCCTACCAATGGCAGCAAGAAATTTGCCATGAATGGAGCTGCTCGGATGTCAGTTT 1290
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1291 ATTCTAAGGTCTCTCCAAACTACTCAACCTCCACCTCCTCGGAACAGCAATGAC 1350
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCAAAAACACTACAGCCCTCCCAAAATAGCCAAAGCTGTCGCCCAAAATTT----- 1404
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTrp 460
QY 1405 -----ACGCAACCACTACACCTCGCAGTAGCAATGAATTTCCGTCACAGACAAA 1455
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgAsnProGln 480
QY 1456 CAACAACCTGCCAGTCTGTATCAGAAATACTACCGTAACCTCCAAATGTACCAAAAGAT 1515
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----GCGTGGCTGCAGTCTTGTCCCTGCTGTG 1548
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1549 GTCATGGTCTCACTACTCTCATTTCTATATTAGTGTGTGCTTGGCACTGG----- 1599

Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1600 -----AGAAACAGAAAAAAGAAAGGACACC-----TATGACTTACTCTACTGG 1647
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1648 GACGGCGCAGTTGGTGGAAAGGAATGAAGCAGTTTCTCTCGCAAGCAGTGGACCAT 1707
Db 559 -----AspIleArgArgPheAspPro----- 565
QY 1708 GAGAAACCCAGTTCCTATAGCAGCAGCAAGTAACTACCTGAGTCCAGAGAAAGTC 1767
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1768 ACCACAGTGTGCGAGTGTACTCTGCAGAGTAT-----GTCAGCCACTGGTAGGAGA 1821
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1822 ATTGTTGGTACACTTCATCAAAAGATCACTTTAAACCA-----GAAGAAGGA 1869
Db 604 LeuGlyProThrValLysSerGluGluThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAGAAGCAGGCTAT-----GCAGACCTAGATCCTTACAACATCA 1908
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspAspLysAspLeuGln-LeuProSerG 643
QY 1909 CCAGGGCAGGAAGTTTATCATGCCT-----ATGCTGAACCACTCCCA 1950
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTrpValTyrAspHisAlaL 663
QY 1951 A---TTACGGGCGCTGAGTATGCAACCCCAATCATCATGTGACACATGTGAGGCAACCCA-- 2005
Db 663 strpLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 2006 -----CAACTTCAG----- 2014
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArg 696
QY 2015 -----TTGGTCAGCCCTCCACATCCACTTCAAGGCTACGGGAACCAA 2058
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 2059 CTTCCCTCCACTAGTGGGAACATTACAATCACTCTCTC----- 2096
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeu 731
QY 2097 -----CAGGACTGACAGCTGCTC-----CTCAGCCAGCCCGCAG 2130
Db 731 lnValValArgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2131 TATGATACCCGGAAGCTGGGAAGCCAGGTCTACTGCCCCAGACGAATTTGTTACCCAG 2190
Db 751 GlySerGluTrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2191 GTGCCACAGACACACAAGATATCAGGACGAGGATGGGAATGTGATGTT 2247
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGluIleSerIle 786
RESULT 5
US-08-936-135-10
; Sequence 10, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH

Db 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 535 TCCAAATTTTGGACCTGAGTTCAGTAAGTACTCCAGCTCGGTGTCGTCTCTCTTT 594
Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
QY 595 GCTGAGATATCTGGAACAATCTCATGATATAGATATCTCGCCATGTCATGCTG 654
Db 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 655 GTGTGCAATGAGAGAGTGTCAACACAGTTGGCGGCCAAATCAGT----- 702
Db 182 LysProArgMetGluIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu 201
QY 703 -----GTTGTAATTAAGTAAAGGTAATCC--- 726
Db 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis 221
QY 727 -----TATTATGAAGTTCTTTGGCTAAACAGTCCACATCTGTG 765
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSerSer 241
QY 766 GTGGACACTTATCAAGTCTTTTACATTAAGACAAGT----- 807
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 808 -----GGATGTTAT 816
Db 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnProGluAsnPheGlnCysAsn 278
QY 817 GGAACACTGGGATGGAGTCTGGTGTGATCGGGATCTCAATAACAGCATCATCTGTG 876
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer--- 297
QY 877 CTGGAGTGGAGTGCACACAGGCAAGACAGTGTGGAACCCAAACCCAGGCTG 936
Db 298 -----ThrPheSerAspGlyArgTyrProGlnGlnSerArgLeu 311
QY 937 AAAAAACCTGGACCGCTGGCTGCTTTGGCACTGATGATACAGTGGTTACAATA 996
Db 312 HisGlyAspAsnGlyTrpThrProAsnLeuAspSerAsnLysGluTrpLeuGlnVal 331
QY 997 GATTTGAATAAGAAAGAAATAACAGCATTAACACCTGGA-----TCCACCATG 1050
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGGACACAATTACTATGTCTGCACAGAATCTGTACAGTATGATGGGCAAGAA 1110
Db 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1111 TGGACTGTGTACAGAGAGCTGTGTGGAGCAAGATAGATATTTCAAGGAACAAGAT 1170
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsp 389
QY 1171 TATCACAGGATGTGGTAAATTAATTTTGGCCACCAATATTGCAAGTTTATTAGATG 1230
Db 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1231 AATCTTACCAATGGCAGCAGAAATTCATGAAATGGAGCTGCTCGGATGTCAGTTT 1290
Db 410 ArgProGlnThrTriPheHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys--- 427
QY 1291 ATTCTAAAGTGTGCTCCCAAACTACTCAACTCCACTCCCTCGGNACAGCAATGAC 1350
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCMAAAACACTACAGCCCTCCAAAATAGCAAGGTGTCGCCCAAAATTT----- 1404
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTrpLeuTrp 460
QY 1405 -----ACGCAACCACTACAACCTCCAGTAGCAATGAATTTCTGCACAGACAA 1455
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgAsnProGln 480

RESULT 7

US-08-936-135-14

; Sequence 14, Application US/08936135

; Patent No. 6054293

QY 1456 CAACAACCTGCAGTCTGATATCAGAAATACTACCGTAACCTCCAAATGTACCAAAAGAT 1515
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----GGCTGGCTGCAGTCTTGTGTCTGCTGTGCTG 1548
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1549 GTCATGGTCTCCTACTCTCTCATTTAGTGTGTGCTTGGCACTGG----- 1599
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTrpIleGlnAsp 540
QY 1600 -----AGAAACAGAAAGAAATAACTGAAGCACCC-----TATGACTTACTTACTG 1647
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro--- 558
QY 1648 GACCGGGCAGGTGTGGTGAAGGAATGAAGCAAGTTCTTCTCTGCAAAAGCACTGGACCAT 1707
Db 559 -----AspIleArgArgPheAspPro----- 565
QY 1708 GAGAAACCCAGTTCGCTATAGCAGCAGCGAAGTTATCACTGCTGATCCACAGAGAGTC 1767
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1768 ACCACAGTCTGCAGGCTGACTCTGCAGAGTAT-----GCTCAGCAGCTGTAGGAGGA 1821
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1822 ATTGTGTGTACTACTCATCAAGATCTACTTTTAACCA-----GAAGAAGGA 1869
Db 604 LeuGlyProThrValLysSerGluGluThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAAGAAGCAGGCTAT-----GCAGACTAGATCTTCACTCA 1908
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspAspLysAspLeuGln-LeuProSerG 643
QY 1909 CCAGGGCAGGAGTTATCATGCT-----ATGCTCAACCACTCCCA 1950
Db 643 yPheAsnCysAsnPheAspPheProGluThrCysGlyTrpValTyrAspHisAlaL 663
QY 1951 A-----TTAGGGGCGCTGAGTATGCAACCCCAATCATCATGTCAGGCGCACCCCA-- 2005
Db 663 strPheLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 2006 -----CAACTTCAG----- 2014
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArg 696
QY 2015 -----TTGGTCAGCCCTCCACATCCACTTCAAGGCTACGGGGAACCAA 2058
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 2059 CTTCCCTCCACTAGTGGAACTTACAATACACTTCTCTC----- 2096
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeu 731
QY 2097 -----CAGGACTGACACTGCTC-----CTCAGCCAGGCCAG 2130
Db 731 InValValArgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2131 TATGATACCCGAAAGTGGGAAGCAGGTCTACCTGCCCGCCAGCAATTTGTTGTACCAG 2190
Db 751 GlySerGluTrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTrpGln 769
QY 2191 GTGCACACAGACACAGAAGTATCAGGAGCAGGAGGATGGGAA 2238
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu 783

QY 1240 CAATGGCAGCAAAATTTGCCATGAATGAGCTGCTCGATGTCAGTTATTCCTAAA 1299
Db 413 ThrTrpHisSerGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1300 GGTGCTGCTCCAAACTTACTCAACCTCCACCTCCCTCGGACAGCATGACCTCAAAAC 1359
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMetLeuSerGly 443
QY 1360 ACTACAGCCCTCCAAAAATGAGCAAGGTCGTCGCCCAAAATTT----- 1404
Db 444 LeuIleAlaSerGlnIleSerAlaSerThrGlnGluTrpSerProSer 463
QY 1405 ACGCAACCACTACAACTCGCAGTAGCAATTTCTCGACAGACAGCAACAACAACT 1464
Db 464 AlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgIleProGlnAlaGluPro 483
QY 1465 GCGAGTCTGATTCAGAAATACCTACGTAACCTCAATGTAAACCAAGATGTA----- 1518
Db 484 GlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGlyValIle 503
QY 1519 -----GCGCTGGCTCCAGTCTTCTGCTCCCTGCTGCTCATGGTC 1557
Db 504 GlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPheValArgLys 523
QY 1558 CTCACACTCTCATCTCATATTAGTGTGCTGTCGCACTGG-----AGA 1602
Db 524 PheLysValSerTySerLeuAsnGlyLysAspTrpGluTrpIleGlnAspProArgThr 543
QY 1603 AACAGAAAGAAATACTGAAGCACC-----TATGACTTACCTTACTGGACCGGCA 1656
Db 544 GlnGlnProLysLeuPheGluGlyAsnMetHisTyAspThrPro----- 558
QY 1657 GGTGGTGGAAAGAAAGCAAGTCTTCTGCTCAAAAGCAGTGGACCATGAGAAACC 1716
Db 559 -----AspIleArgPheAspPro-----Ile 566
QY 1717 CCAGTTCGTATGACAGCAGCAAGTAAATCACCTGAGTCCAAGAGAGTCAACACAGTG 1776
Db 567 ProAlaGlnTrpValArgValTyProGluArgTrpSerProAlaGlyIleGlyMetArg 586
QY 1777 CTGACGCTGACTCGCAGATAT-----GCTCAGCCACTGGTAGGAAATGTGGT 1830
Db 587 LeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValLysThrLeuGlyPro 606
QY 1831 ACATTCTCAAGATCTACTTTAAACCA-----GAAGAGGAAGAAGCA 1878
Db 607 ThrValLysSerGluGluThrThrThrProTyProThrGluGluAlaThrGluCys 626
QY 1879 GGC-----TATGCAGACTAGATCTTACAACCTCACAGGCGCAGAAC-- 1921
Db 627 GlyLysAsnCysSerPheGluAspAspLysAsp-LeuGlnLeuProSerGlyPheAsnCy 646
QY 1922 -----TTTATCATGCGCT-----ATGCTGAACCACTCCCAA--TTACG 1956
Db 646 sAsnPheAspPheLeuGluGluProCysGlyTrpMetTyAspHisAlaLysTrpLeuArg 666
QY 1957 GGGCTGATGATGACCAACCCCAATCATCATGACATGTGAGGACCCCAACAA----- 2008
Db 666 g-----ThrThrTrpAlaSerSerSerProAsnAspArgTh 679
QY 2008 ----- 2008
Db 679 rPheProAspAspArgAsnPheLeuArgLeuGlnSerAspSerGlnArgGluGlyGlnTy 699
QY 2009 -CTTCAGTTGTGACGCTCCACATCCACTTTCAAGCTAGCGGAACACACCTCCCA 2067
Db 699 rAlaArgLeuIleSerProPro-ValHisLeu-----ProArgSerProV 714
QY 2068 CTAGTGGGAACCTACATACACTCTCTC----- 2096
Db 714 aCysMetGluPheGlnTrpGlnAlaThrGlyArgGlyValAlaLeuGlnValVala 734
QY 2097 -----CAGGACTGACAGCTGCTC-----CTCAGCCCGAGCCGATGATACC 2139

Db 734 rgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlyGlu 753
QY 2140 CCGAAACCTGGAGCCAGCTTACCTGCCCGACAGCAATTTGGTGACAGGTCACACAG 2199
Db 754 TrpLysHisGlyArgIleLeuLeuProSerTyAsp---MetGluTrpGlnIle----- 770
QY 2200 AGCACACAAGAGTATCAGGACGAGGAGGATGGGAA 2238
Db 771 ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu 783
RESULT 10
US-08-936-135-20
; Sequence 20, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 926 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-936-135-20
Alignment Scores:
Pred. No.: 2,59e-28 Length: 926
Score: 398.50 Matches: 199
Percent Similarity: 36.37% Conservative: 112
Best Local Similarity: 23.27% Mismatches: 340
Query Match: 9.62% Indels: 204
DB: 3 Gaps: 36

US-10-060-830-1 (1-2280) x US-08-936-135-20 (1-926)
QY 76 ATGCTCTGTCTCTCTG-----CTCTTACTTGTCTCTCTCTGCTGCTGAGGACGCT 129
Db 1 MetAspMetPheProLeuLeuTrpValPheLeuAlaLeuTyPheSerArgHisGlnVal 20
QY 130 GGAGCCCAAGAGGATGATGTGACACACTAGGCTGAGGCTGAGAGTGGAAACCCCT 189
Db 21 ArgGlyGlnProAspProCysGlyArgLeuAsnSerLysAspAlaGlyTyrlle 40
QY 190 ACATCCATAAATACCCACAGACCTATCCCAACAGCACTGTTGTGAATGGAGATCCGT 249
Db 190 ACATCCATAAATACCCACAGACCTATCCCAACAGCACTGTTGTGAATGGAGATCCGT 249


```
Db 296 -----GlnTyrGlyThrAsnTrpSerValGluArgSerArgLeuAsnTyrProGlu 312
Qy 949 CCGCTTGGGCTGCTTTGGCCACGATGATACCAATACCAAGTGTACAAATAGATTGTAATAAG 1008
Db 313 AsnGlyTrpThrProGlyGluAspSerTyrArgGluTrpIleGlnValAspLeuGlyLeu 332
Qy 1009 GAAAGAAAATPACAGGCATTATACCACTGGA-----TCCACCATGGTGGAGCACAAT 1062
Db 333 LeuArgPheValThrAlaValGlyThrGlnGlyAlaIleSerLysGluThrLysLysLys 352
Qy 1063 TACTATGTCTGCTGCTACAGATCCTGTACAGTGTATGATGATGATGATGATGATGATGATGAT 1122
Db 353 TyrTrpValLysThrTyrArgValAspIleSerAsnGlyGluAspTrpIleThrLeu 372
Qy 1123 AGAGAGCCTGTGTGGAGCAAGATAATTTCAAGGAAACAAAGATTATCACCAGGAT 1182
Db 373 LysGluGly-----AsnLysAlaIleIlePheGlnGlyAsnThrAsnProThrAspVal 390
Qy 1183 GTGCGTAATAACTTTTCCCAACCAATATTACAGCTTTTATTATAGAGTGAATCCTACCCAA 1242
Db 391 ValPheGlyValPheProLysProLeuIleThrArgPheValArgIleLysProAlaSer 410
Qy 1243 TGGCAGCAAAATGCGCATGAAATGAGCTGCTCGCATGCTCAGTATTATTCCTAAAGGT 1302
Db 411 TrpGluThrGlyIleSerMetArgPheGluValTyrGlyCys----- 424
Qy 1303 CGTCTCCAAAACCTTACTCAACCTCCA 1329
Db 425 -----LysIleThrAspTyrPro 430

RESULT 13
US-08-480-229C-10
; Sequence 10, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0026-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-229C-10
```

```
Alignment Scores: 2.96e-25 Length: 480
Pred. No.: 364.00 Matches: 138
Score: 37.12% Conservative: 58
Percent Similarity: 26.14% Mismatches: 170
Best Local Similarity: 8.79% Indels: 162
Query Match: 2 Gaps: 21
DB: 21

US-10-060-830-1 (1-2280) x US-08-480-229C-10 (1-480)
Qy 13 CCCTGGGCGGCTCCCTCCCTCTCCGCTCCCTCCCT----- 48
Db 7 AlaTrpLeuValGlyLeuSerLeuGlyValProGlnPheGlyLysGlyAspIleCys 26
Qy 49 -----CCCTGTCTCCAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTA 99
Db 27 AsnProAsnProCysGluAsnGlyGly----- 35
Qy 100 CTGTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 159
Db 36 IleCysLeuSerGlyLeuAlaAspSerPheSerCysGlu-----CysProGlu 52
Qy 160 ACTGTACTAGCCCTGAGAGTGGACCTTACATCCATAACCTACCCACACACCTATCCC 219
Db 53 GlyPheAlaGlyProAsnCysSerSerValValGluValAlaSerAspGluGluLysPro 72
Qy 220 -----AACAGCAGCTGTTTGTGAATGGGAG 243
Db 73 ThrSerAlaGlyProCysIleProAsnProCysHisAsnGlyGlyThrCysGlu----- 90
Qy 244 ATCCGTGTAAGAGGGAGAGAGTTCGCATCAAAATTTGTGACTTTGACATTTGACATTTGAAGAT 303
Db 91 -----IleSerGluAlaTyrArg-----GlyAspThrPheIleGlyTyr 103
Qy 304 TCTGATTTCTGT-----CACTTTAATTACTTGAGAAATTTAATGGAATGGAGTCCAGC 357
Db 104 ValCysLysCysProArgGlyPheAsnGlyIleHisCysGlnHisAsnIleAsnGluCys 123
Qy 358 AGAAGTGAATAGGCAAAATCTGTGCTGT-----GGTTTGCATAATGAACATTTCAAT 411
Db 124 GluAlaGluProCysArgAsnGlyGlyIleCysThrAspLeuValAlaAsnTyrSerCys 143
Qy 412 GAATCAAAAGCAATGAAATCATCATGTGTTTCATGAGT----- 450
Db 144 GluCysProGlyGlu-----PheMetGlyArgAsnCysGlnTyrLysCys 158
Qy 451 -----GGAATCCAT-----GTTTCTGGACGGGATTTTGGCCTCA--- 486
Db 159 SerGlyHisLeuGlyIleGluGlyIleIleSerAsnGlnGlnIleThrAlaSerSer 178
Qy 487 -----TACTCTCTTATAGATAA 504
Db 179 AsnHisArgAlaLeuPheGlyLeuGlnLysTyrTyrProTyrTyrAlaArgLeuAsnLys 198
Qy 505 CAAGATCTTAATTTACTTTTGGACACTGCATCCAAT----- 540
Db 199 LysGlyLeuIleAsnAlaTrpThrAlaAlaGluAsnAspArgTyrProTrpIleGlnIle 218
Qy 541 -----TTTTTG 546
Db 219 AsnLeuGlnArgLysMetArgValThrGlyValIleThrGlnGlyAlaLysArgIleGly 238
Qy 547 GAACCTGATTCAGTAAG-----TACTGCCAGCGGTGTCTGCTCTCTCT 591
Db 239 SerProGluTyrIleLysSerTyrLysIleAlaTyrSerAsnAspGlyLysThrTrpAla 258
Qy 592 TTTGCTGAGATATCTGGAACAATTCCTCATGGATATAGAGATTCTCGCATTTGTCATG 651
Db 259 MetTyrLysValLysGlyThr-----AsnGluGluMetValPhe 271
Qy 652 GCTGTGTGTCATGCAGGAGTAGTGTCAAAACAGTTGGCGGCCCAATCATGTTGTGAAT 711
Db 711 -----TTTTTT
```

Db	272	ArgGly	-----AsnValAspAsnAsnThr-----	279
QY	712	AGTAAAGGATT	CCCTATTATGAAAGTTCTTTGGCTAACAAACGTCACATCTGTGGTGGGA	771
Db	280	-----ProTyr-----		290
QY	772	-----	-----AlaAsnSerPheThrProProIleLys	786
Db	291	AlaGlnTyrValArgLeuTyrProGlnIleCysArgArgHisCysThrLeuArgMetGlu	310	
QY	787	CTTTTTCATATTAAAGCAAGTGGATGTTATGGAACACATGGGGAGTGAGTCTGGGTGCATC	846	
Db	311	LeuLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMetLysSerGlyHisIle	330	
QY	847	GCGGATCCTCAATAACACGATCATCTGCTGCTGGAGTGGACTGACACACAGGGCAAGAG	906	
Db	331	GlnAspTyrGlnIleThrAlaSerValPheArgThrLeuAsnMet-----	348	
QY	907	ACAGTTGGAACCCAAAAAAGCCAGGCTGAAAAAACCCTGGACCG-----	960	
Db	349	PheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysValAsnAlaTrpThr	368	
QY	961	CGTTTGGCCACTGATGAATACCACTGGTGTACAAATAGATTTGAATAGGAAAAAGAAATA	1020	
Db	369	SerGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuLeuValProThrLysVal	388	
QY	1021	ACAGGCATTTAAACCACTGGATCCACCATGGTGGAGCACAATTAATATGTGTCTGCCTAC	1080	
Db	389	ThrGlyIleIleThrGlnGlyAlaLysAspPheGlyHisValGlnPheValGlySerTyr	408	
QY	1081	AGAATCCTCTACAGTATGATGGCAGAAAATGGACTGTGTACAGAGCCCTGGTGTGGAG	1140	
Db	409	LysLeuAlaTyrSerAsnAspGlyGluHisTrpMetValHisGlnAspGluLysGlnArg	428	
QY	1141	CAGATAAGATATTTCAAGGAACAAGATATCATCCAGGATGTCGCTAATACTTTTG	1200	
Db	429	LysAspLysValPheGlnGlyAsnPheAspAsnAspThrHisArgLysAsnValIleAsp	448	
QY	1201	CCACCAATTTCCAGGTTTTATTAGATGAATCCTCCCAATGGCAGCAGAAAATTGCC	1260	
Db	449	ProProIleTyrAlaArgPheIleArgIleLeuProTrpSerTyrGlyArgIleThr	468	
QY	1261	ATGAAATAGGAGTGTGTCGGATGT	1284	
Db	469	LeuArgSerGluLeuLeuGlyCys	476	

RESULT 14

US-08-659-235C-10
; Sequence 10, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quertemous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zuparcic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/659,235C
; FILING DATE: 05-JUN-1996

Db 199 LysGlyLeuIleAsnAlaTrpThrAlaAlaGluAsnAspArgTrpProTrpIleGlnIle 218
Qy 541 -----TTTTG 546
Db 219 AsnLeuGlnArgLysMetArgValThrGlyValIleThrGlnGlyAlaLysArgIleGly 238
Qy 547 GAACCTGAGTTCAGTAAG-----TACTGCCACAGCTGGTTCCTGCTTCCT 591
Db 239 SerProGluTrpIleLysSerTrpLysIleAlaLysSerAsnAspGlyLysThrTrpAla 258
Qy 592 TTTGCTGAGATCTCGAAACAACTCCTCATGATATAGATGATCGCCATCTGTCATG 651
Db 259 MetTrpLysValLysGlyThr-----AsnGluGluMetValPhe 271
Qy 652 GCTGGTGCATCAGCAGAGTAGTGTCAACACAGTGTGGCGGCCCAATCAGTGTGTAAT 711
Db 272 ArgGly-----AsnValAspAsnAsnThr----- 279
Qy 712 AGTAAGGATTCCTTATATGAAGTCTTTGGTCAACACAGTCACATCTGTGGTGGGA 771
Db 280 -----ProTyr-----AlaAsnSerPheThrProProIleLys 290
Qy 772 -----CACTTATCTACAAGT 786
Db 291 AlaGlnTyrValArgLeuTyrProGlnIleCysArgArgHisCysThrLeuArgMetGlu 310
Qy 787 CTTTTCATATTAAGCAAGTGTATGGAACACTGGGATGGAGTGTGTGTGATC 846
Db 311 LeuLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMetLysSerGlyHisIle 330
Qy 847 GCGATCCTCAAAATACAGCATCATCTGCTGGAGTGGACTGACACACAGGCGCAAG 906
Db 331 GlnAspTyrGlnIleThrAlaSerSerValPheArgThrLeuAsnMet-----AspMet 348
Qy 907 AACAGTGGAAACCCAAAAGCAGCCTGAAAAAACCTGGACCG-----CCTTGGGCT 960
Db 349 PheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysValAsnAlaTrpThr 368
Qy 961 GCTTTTCCACTGATGAATACCACTGCTGTCAAAATAGATTTGAATAGGAAAGAAATA 1020
Db 369 SerGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuValProThrLysVal 388
Qy 1021 ACAGGCATTATAACCACTGATCCACCATGCTGGAGCAGCATTTACTATGTCTGCTGCCTAC 1080
Db 389 ThrGlyIleIleThrGlnGlyAlaLysAspPheGlyHisValGlnPheValGlySerTyr 408
Qy 1081 AGAATCCTGACATGATGATGGCAGAAATGGAGTGTGTACAGAGACCTGGTGGAG 1140
Db 409 LysLeuAlaTyrSerAsnAspGlyGluHisTrpMetValHisGlnAspGluLysGlnArg 428
Qy 1141 CAAGATAAGATATTTCAAGGAACAAGATATACCCAGGATGTCGGTAATAACTTTTGG 1200
Db 429 LysAspLysValPheGlnGlyAsnPheAspAsnAspThrHisArgLysAsnValIleAsp 448
Qy 1201 CCACCAATTATTGACGCTTTATTAGATGAATCCTCCCAATGGCAGCAGAAATGCC 1260
Db 449 ProProIleTyrAlaArgPheIleArgIleLeuProTrpSerTrpTyrGlyArgIleThr 468
Qy 1261 ATGAAATGGAGCTGCTGGATG 1284
Db 469 LeuArgSerGluLeuLeuGlyCys 476

RESULT 15
US-08-480-229C-14

; Sequence 14, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quertemous, Thomas
; APPLICANT: Hogan, Bridg
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; CELL LOCUS-1

NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,229C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-229C-14

Alignment Scores:
Pred. No.: 1.87e-24 Length: 513
Score: 355.50 Matches: 137
Percent Similarity: 38.45% Conservative: 51
Best Local Similarity: 28.02% Mismatches: 180
Query Match: 8.58% Indels: 121
DB: 2 Gaps: 20

US-10-060-830-1 (1-2280) x US-08-480-229C-14 (1-513)

Qy 68 CTTCTCTCATCG-----CTCTGT-----TCTCTCTGCTCTTACTTGTCC 106
Db 60 ProAsnProCysGluAsnGlyGlyLeuProGlyLeuAlaValGlySerPheSer 79
Qy 107 TGCTCTCTGC-----TGCTCGAGGACGCTGGAGCCGAGCAAG 142
Db 80 CysGluCysProAspGlyPheThrAspProAsnCysSerValValGluValAla--S 99
Qy 143 GTGATGATGTGGACACACTGTACTAGCCCTGAGATGGAGACCTTACATCCATAACT 202
Db 99 erAspGluGluGluProThrSerAlaGlyProCysThr----- 111
Qy 203 ACCCAGACACCTATCCCAACAGCAGTGTGTGTGATGGAGATCCGTGTAAGATGGAG 262
Db 112 --ProAsnProCysHisAsnGlyGlyThrCysGlu-----IleSerG 125
Qy 263 AGAGAGTTTCGCATCAAAATTTGGTGACTTTGACATTTGAATGATTCGATCTTGT-----C 316
Db 125 luAlaTyrArg-----GlyAspThrPheIleGlyTyrValCysLysCysProArg 142
Qy 317 ACTTTAATTACTTGAGATTTATTAATGAATTTGAGTGCAGCACTGAATAGGCAAT 376
Db 142 lyPheAsnGlyIleHisCysGlnHisAsnIleAsnGluCysGluValGluProCysLys 162
Qy 377 ACTGTGGTCTG-----GGTTGCAAAATGAACCTTCAATTAATCAAAAGCAATGAA 430
Db 162 snGlyGlyIleCysThrAspLeuValAlaAsnTyrSerCysGluCysProGlyGlu---- 180
Qy 431 TCACATTGCTGTTCATGAGT-----GGAATCC 457

Job time : 46.2425 secs

```
Db 181 -----PheMetGlyArgAsnCysGlnTyrLysCysSerGlyProLeuGlyIleG 197      |||||
QY 458 AT-----GTTTCTGGACCGGATTTTGGCCTCA----- 486      |||||
Db 197 luGlyGlyIleSerAsnGlnGlnIleThrAlaSerSerThrHisArgAlaLeuPheG 217      |||||
QY 487 -----TACTCTGTATAGATAAACAAGATCTAATTACTTGT 523      |||||
Db 217 lyLeuGlnLysTrpTyrProTyrTyrAlaArgLeuAsnLysGlyLeuIleAsnAlaT 237      |||||
QY 524 TGGACACTGCATCCAAATTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGGTCTC 583      |||||
Db 237 rpThrAlaAlaGluAsn-----AspArgTrpLysArgTrpIleGlnIleAsn--L 253      |||||
QY 584 TGCTTCTCTTTCCTCAGATATCTGCACAACATTCCTCATGGATATAGA----- 630      |||||
Db 253 euGlnArgLysMetArgValThrGlyValIleThrGlnGlyAlaLysArgIleGlySerP 273      |||||
QY 631 -----GATTCTCGCCATTGTGCATGG 652      |||||
Db 273 roGluTyrIleLysPheTyrLysIleAlaTyrSerAsnAspGlyLysThrTrpAlaMetT 293      |||||
QY 653 CTGGTGTGCATCCAGGAGTAGTGTCAACACAGCTTGGCGCGCCAAATCAGTGTGTGAATTA 712      |||||
Db 293 yrLysValLysGlyThrAsnGluAspMetValPheArgGlyAsn-----IleA 309      |||||
QY 713 GFAAAGGTATTCCTTATTAAGAACTTCTTGGCTCAACAGCTCACATCTCTGGTGGGA- 771      |||||
Db 309 spAsnAsnThrProTyr-----AlaAsnSerPheThrProIleLysA 324      |||||
QY 772 -----CACTTATCTACAAGTC 787      |||||
Db 324 laGlnTyrValArgLeuTyrProGlnValCysArgArgHisCysThrLeuArgMetGluL 344      |||||
QY 788 TTTTACATTTAAGACAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATCG 847      |||||
Db 344 euLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMetLysSerGlyHisIleG 364      |||||
QY 848 CGGATCCTCAATACACATCATCTGTCTGGAGTGACTGACCACACAGGCGCAAGAGA 907      |||||
Db 364 lnAspTyrGlnIleThrAlaSerIlePheArgThrLeuAsnMet-----AspMetP 382      |||||
QY 908 ACAGTTGGAACCCAAAAGCCAGCTGAAAACACCTGGACCG-----CCTTGGGCTG 961      |||||
Db 382 heThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysValAsnAlaTrpThrS 402      |||||
QY 962 CTTTGGCCACTGATGAATACCACTGGTTACAAATAGATTTTGAATAAGGAAAAGAAATAA 1021      |||||
Db 402 erGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuLeuValProThrLysValT 422      |||||
QY 1022 CAGGCATTATACCACTGGATCCACCCTGGTGGAGCACAAATTAATGTGTCTGCCTACA 1081      |||||
Db 422 hrGlyIleIleThrGlnGlyAlaLysAspPheGlyHisValGlnPheValGlySertyrL 442      |||||
QY 1082 GAATCCTCTACAGTATGGCAGAAATGGACTGTGTACAGAGAGCCTGGTGTGGAGC 1141      |||||
Db 442 ysLeuAlaTyrSerAsnAspGlyGluHisTrpThrValTrpGlnAspGluLysGlnArgL 462      |||||
QY 1142 AAGATPAGATATTTCAAGGAACCAAGATTTATCACAGGATGTGCTAATAACTTTTTCG 1201      |||||
Db 462 ysAspLysValPheGlnGlyAsnPheAspAsnAspThrHisArgLysAsnValIleAspP 482      |||||
QY 1202 CACCAATTTATTCAGCTTTTATAGAGTGAATCCTACCAATGGCAGCAGAAATTCGCA 1261      |||||
Db 482 roProIleTyrAlaArgHisIleIleLeuProTrpSerTrpTyrGlyArgIleThrL 502      |||||
QY 1262 TGAATAATGAGCTGCTCGGATCT 1284      |||||
Db 502 euAlaSerGluLeuLeuGlyCys 509      |||||
```

Search completed: January 21, 2003, 09:50:50

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 21, 2003, 02:16:00 ; Search time 54.9883 Seconds
(without alignments)
10613.845 Million cell updates/sec

Title: US-10-060-830-2
Perfect score: 3969
Sequence: 1 atgcctctgtctctctgtct.....ttttaagaatacctttga 2190

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delopt 6.0, Delopt 7.0

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cpr2_1/USPTO.spool_QM10060830/runat_16012003_091238_26833/app_query.fasta_1.6933
-DB=A_Geneseq_101002 -QPM=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdd
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830_@CGN_1.1.230_@runat_16012003_091238_26833 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	3572	90.0	669	21	AAV70539	Human Factor 8 Hom
2	2806	70.7	583	22	AAV75450	Human Colon cancer
3	1841.5	46.4	385	23	AAE22716	Human neuropilin-H
4	1841.5	46.4	385	23	AAU79460	Human neuropilin-H
5	1841.5	46.4	385	23	ABB97386	Novel human protei
6	1750.5	44.1	365	23	AAE22721	Human neuropilin-H
7	1042	26.3	715	22	AAU00670	Human TANGO 229 po
8	1039	26.2	137	22	ABG01298	Novel human diagno
9	924.5	23.3	539	22	AAU00630	Novel human protei
10	924.5	23.3	586	22	AAU00629	Novel human protei
11	860.5	21.7	487	22	AAU00628	Novel human protei
12	574.5	14.5	503	21	AA19126	Polypeptide isolat
13	540.5	13.6	398	23	AAE22715	Human neuropilin-H
14	540.5	13.6	398	23	AAU79459	Human neuropilin-H
15	421.5	10.6	925	20	AAW96308	Neuropilin-2, Rat
16	414.5	10.4	901	20	AAW96356	Mouse semaphorin r
17	414.5	10.4	906	20	AAW96257	Mouse semaphorin r
18	414.5	10.4	909	20	AAW96249	Rat semaphorin rec
19	412.5	10.4	909	20	AAW96250	Mouse semaphorin r
20	412.5	10.4	909	21	AAE24214	Mouse soluble neur
21	412.5	10.4	914	20	AAW96251	Mouse semaphorin r
22	412.5	10.4	926	20	AAW96252	Mouse semaphorin r
23	412.5	10.4	931	20	AAW96253	Mouse semaphorin r
24	412	10.4	75	22	AB42217	Peptide #9723 enco
25	412	10.4	75	22	AAW63102	Human brain expres
26	412	10.4	75	22	AAW75913	Human bone marrow
27	412	10.4	75	22	AAW36024	Peptide #10061 enc
28	412	10.4	75	23	AB45349	Human peptide enco
29	408.5	10.3	889	21	AAE24216	Soluble neuropilli
30	404.5	10.2	957	22	AAW25770	Human protein sequ
31	400.5	10.1	538	22	AAU02950	Angiotensin conver
32	400.5	10.1	600	22	AAU02948	Angiotensin conver
33	400.5	10.1	644	20	AAU06319	Human soluble neur
34	400.5	10.1	840	22	AAU02949	Angiotensin conver
35	400.5	10.1	856	22	AAE62478	Human NP-1 recepto
36	400.5	10.1	923	20	AAW06317	Human neuropilin-1
37	400.5	10.1	923	20	AAV23247	Human VEGF165R/NP
38	400.5	10.1	923	20	AAW96246	Human semaphorin r
39	400.5	10.1	923	23	AAE22717	Human neuropilin-1
40	400.5	10.1	924	22	AAW62476	Human neuropilin-1
41	398.5	10.0	909	20	AAW96254	Mouse semaphorin r
42	398.5	10.0	923	20	AAV14562	Human VEGF(165)R/N
43	398.5	10.0	926	20	AAW96255	Human semaphorin r
44	398.5	10.0	926	23	AAE22718	Human neuropilin-2
45	398.5	10.0	931	20	AAW06318	Human neuropilin-2

ALIGNMENTS

RESULT 1
AAV70539
ID AAV70539 standard; Protein; 669 AA.
XX
AC AAV70539;
XX
DT 04-JUL-2000 (first entry)
XX
DE Human Factor 8 Homologue.
XX
KW Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
cerebroprotective; therapeutic; coagulation related disorder;
KW haemophilia; stroke; screening.
XX
OS Homo sapiens.
XX
PN WO200012532-A1.
XX
PD 09-MAR-2000.

XX 20-AUG-1999; 99WO-US19047.
XX 31-AUG-1998; 98US-0098521.
XX (ELIL) LILLY & CO ELI.
XX Rostock PRJ, Su W, L1 XM;
XX WPI; 2000-256580/22.
XX N-PSDB; AAZ51872.
XX Factor 8 homolog polypeptides and nucleic acids encoding them for
XX treating coagulation related disorders such as hemophilia and stroke
XX Claim 3; Page 64-66; 68pp; English.
XX The present sequence is a human Factor 8 homologue (F8H),
XX a coagulation cofactor which is selectively expressed in
XX haematopoietic, heart and reproductive tissues. It has haemostatic and
XX cerebroprotective activities. The F8H contains a Factor 5/8 signature
XX and is useful as a therapeutic for treating coagulation related diseases
XX such as haemophilia and stroke. The nucleic acid is useful as
XX hybridisation probe and amplification primer for detecting deficiencies
XX in the level of F8H mRNA, for screening F8H gene mutations and for
XX monitoring regulation of gene expression. Fragments of the nucleic acid
XX are also useful as diagnostic probes and primers, and can be used in
XX screening methods such as those using DNA chips. The present sequence is
XX also useful as a target to screen therapeutically useful modulators
XX of the F8H.
XX SQ Sequence 669 AA;

Alignment Scores:

Pred. No.: 0 Length: 669
Score: 3572.00 Matches: 669
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.00% Indels: 0
DB: 21 Gaps: 0

US-10-060-830-2 (1-2190) x AAY70539 (1-669)

QY 181 ATGGAGAGAGAGTCCGATCAATTTGGTGCATTTTGACATTTGAAGATTCCTGATTTGT 240
DB 1 MetGlyGluArgValArgIleGlyPheGlyAspPheAspPheGluAspSerAspSerCys 20
QY 241 CACTTTAATTAATCTGAGAAATTTAATGGAATTTGGAGTACGACAGAACTGAAATAGGCAAA 300
DB 21 HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys 40
QY 301 TACTGTGCTCGGGTTCGAATGACCATTCATTAATGAATCAAAAGGCAATGAATCACA 360
DB 41 TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr 60
QY 361 TTGCTGTCATGAGTGAATCCATGTTCTCGACGCGGATTTTGGCTCATCTCTGTT 420
DB 61 LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal 80
QY 421 ATAGATAACAGATCTAATTAATCTGTTGGACACTGCATCCCAATTTTGGACCTGAG 480
DB 81 IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu 100
QY 481 TTCAGTAGTACTGCCAGCTGGTGTCTGCTGCTTCTTCTGCTGAGATATCTGGACAAAT 540
DB 101 PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIle 120
QY 541 CCTCATGATATAGAGATCTCGCATGTGTCATGGTGGTGTGTCATGCAGGAGTAGTG 600
DB 121 ProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValVal 140
QY 601 TCAACACAGTGGGGCGCAATCAGTGTGTAATTAAGTAAGGATTTCCCTATTATGAA 660
|||||

DB 141 SerAsnThrLeuGlyGlyGlnIleSerValIleSerLysGlyIleProTyrTyrGlu 160
QY 661 AGTTCTTTGGCTAACACAGCTCAGATCTGGTGGGACACTATCTACAAGTCTTTTACA 720
DB 161 SerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr 180
QY 721 TTTAAGACAAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGGTGTGATCGCGATCCT 780
DB 181 PheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspPro 200
QY 781 CAAATACAGCATCATCTGCTGGAGTGGACTGACACACAGGCGAAGCAAGATTGG 840
DB 201 GlnIleThrAlaSerSerValLeuGluThrPheThrAspHisThrGlyGlnGluAsnSerTyr 220
QY 841 AAACCCAAAAAGCCAGGCTGAAAAAACCCTGGACCGCTTGGCTGGCTTTTCCCATCAT 900
DB 221 LysProLysLysAlaArgLeuLysLysProGlyProThrPheAlaPheAlaThrAsp 240
QY 901 GAATACAGTGGTTACAAATAGATTGTAATAGGAAGAAAGAAATAACAGGATTATAACC 960
DB 241 GluTyrGlnThrPheGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleThr 260
QY 961 ACTGGATCCACCATGGTGGACACAATTACTATGCTGTGCTACAGAAATCCTGTACAGT 1020
DB 261 ThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSer 280
QY 1021 GATGATGGGAGAAATGGACTGTGTACAGAGAGCTGGTGGGACCAAGATAAGATATT 1080
DB 281 AspAspGlyGlnLysTyrThrValTyrArgGluProGlyValGluGlnAspLysIlePhe 300
QY 1081 CAAGGAACAAGATTTATCACAGGATGTGGTAAATTAATTTTGGCCACCAATATTGCA 1140
DB 301 GlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProIleIleAla 320
QY 1141 CGTTTATTAGAGTGAATCCCTACCAATGGCAGCAGAAATTTGCCATGAAATGGAGCTG 1200
DB 321 ArgPheIleArgValAsnProThrGlnThrGlnGlnLysIleAlaMetLysMetGluLeu 340
QY 1201 CTCGGATGTCAGTTTATTCCTAAAGGTCGTCTCCAAAACCTTACTCAACCTCCCATCCT 1260
DB 341 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProPro 360
QY 1261 CGGAACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAATAGCCAAAGTGTGCTGCC 1320
DB 361 ArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla 380
QY 1321 CCAAAATTCGCAACCACTACAACTCGCAGTAGTGAATGAAATTTCTCGCACAGACAGAA 1380
DB 381 ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 400
QY 1381 CAACAACCTCCAGCTCCTGATATCAGAAATACCTACCGTAACCTCCAAATGTAAACCAAGAT 1440
DB 401 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 420
QY 1441 GTACGCTGGCTGCGAGTCTTGTCCCTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1500
DB 421 ValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 440
QY 1501 ATATTAGTGTGCTGGCACTGGAGAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1560
DB 441 IleLeuValCysAlaThrPheHisThrArgAsnArgLysLysLysLysLysLysLysLys 460
QY 1561 TTACCTTACTGGGACCGCGAGGTTGGTGGAAAGGAATGAAGCAGTGTCTTCTCTCAAAA 1620
DB 461 LeuProTyrThrAspArgAlaGlyTyrThrPheLysGlyMetLysGlnPheLeuProLys 480
QY 1621 GCAGTGGACCATGAGAAACCCCGATTCGCTATAGCAGCAGCAGGAGTGAATCACCCTGAGT 1680
DB 481 AlaValAspHisGluGluThrProValArgTyrSerSerSerGluValAsnHisLeuSer 500
QY 1681 CCAAGAGAAGTCCACACAGTGTGCTGAGCTGACTCTGCAGAGTAGTGTGCTGAGCCTGTTA 1740
DB 501 ProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuVal 520
|||||

Wed Jan 22 14:57:59 2003

us-10-060-830-2.n2p.ra1

Page

Search completed: January 21, 2003, 09:51:21
Job time : 45.6409 secs

QY 1993 CTAGTGGGAACCTACATACACTCTCTC----- 2021
Db 714 aLCysMetGluPheGlnThrGlyGlyArgGlyValAlaLeuGlnValVala 734
QY 2022 -----CAGGACTGACAGCTGCTC-----CTCAGCCAGGCCAGCTATGATACC 2064
Db 734 rgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlyGlu 753
QY 2065 CGAAAGCTGGGAAGCAGGCTACCTGCGCCAGCAGCAATTTGGTACCAGGTGCCACAG 2124
Db 754 TrpLysHisGlyArgIleLeuProSerTyrAsp---MetGluTyrGlnIle----- 770
QY 2125 AGCACACAAGAATATCAGGACGAGGAGGATGGGAA 2163
Db 771 ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu 783
RESULT 11
US-08-936-135-6
; Sequence 6, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 923 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-936-135-6
Alignment Scores:
Pred. No.: 1.64e-28 Length: 923
Score: 391.50 Matches: 129
Percent Similarity: 40.04% Conservative: 60
Best Local Similarity: 27.33% Mismatches: 176
Query Match: 9.86% Indels: 107
DB: 3 Gaps: 18
US-10-060-830-2 (1-2190) x US-08-936-135-6 (1-923)
QY 16 CTGCTCTTACTGCTCTCTGCTCGAGGACGCTGGAGCCAGCAAGGTGATGGA 75
Db 7 LeuLeuCysAlaThrLeuAlaLeuAlaLeuAlaLeuAlaGlyAlaPheArgSerAspLys 26
QY 76 TGTGGACACACTGTACTAGGCCCTCAGAGTGGGAACCCCTTACATCCATAAACCACACAG 135

Db 27 CysGlyGlyThrIleLysIleGluAsnProGlyTyrLeuThrSerProGlyTyrProHis 46
QY 136 ACCATCCCAACACGACACTGTTGTGAATGGAGATCCGTGTA---AAGATGGGAGAGAGA 192
Db 47 SerTyrHisProSerGluLysCysGluTrpLeuIleGlnAlaProGluProTyrGlnArg 66
QY 193 GTTCGCATCAAAATTGGT---GACTTTGACATTGAAGATTCTGTCTCTGCTTCAAT 249
Db 67 IleIleIleAsnProHisPheAspLeuGluAspArgAsp---CysLysTyrAsp 85
QY 250 TACTGTAGAATTATATGAATTGGATTGGACTCAGCAGAACTGAATAGGCAATATCTGCTG 309
Db 86 TyrValGluValIleAspGlyLysGluGlyArgLeuTrpGlyLysPheCysGly 105
QY 310 CTGGGGTTGCAAAATGAACCATTTCA---ATTGAATCAAAAGGCAATGAATCACATTCTG 366
Db 106 -----LysIleAlaProSerProValValSerSerGlyProPheLeuPheIleLys 122
QY 367 TTCATGAGTGAATCCATGTTTCTGGACGCGGATTTTGGCTCATATCTGTATATAGAT 426
Db 123 PheValSerAspTyrGluThrHisGlyAlaGlyPheSerIleArgTyrGluIlePheLys 142
QY 427 AAACAAGATCTAATTACTTGTGGACACTGCATCCCAATTTTGGGAACCTCAGTTCAGT 486
Db 143 ArgGlyPro-----GluCysSerGlnAsnTyrThrAlaPro----- 154
QY 487 AAGTACTGCCCGCTGGTGTCTGCTCTTCTGCTGAGATATCTGGAACAATTCCTCAT 546
Db 155 -----ThrGlyValIleLysSer 160
QY 547 ---GGATATAGAGATTCCTCGCCATGTCATG-----GCTGGT 582
Db 161 ProGlyPheProGluLysTyrProAsnCysLeuGluCysThrTyrIleIlePheAlaPro 180
QY 583 GTGCATGCAGGAGTAGTG-----TCAAACACG 609
Db 181 LysMetSerGluIleIleLeuGluPheGluSerPheAspLeuGluGlnAspSerAsnPro 200
QY 610 TTGGCGCGCCAAATC-----AGTGTGTAATAGTAAAGGTATCC----- 651
Db 201 ProGlyGlyMetPheCysArgTyrAspArgLeuGluIleTrpAspGlyPheProGluVal 220
QY 652 -----TATTATGAAAGTCTTTGGCTAACACACGCTCAGTCTGTGTG 693
Db 221 GlyProHisIleGlyArgTyrCysGlyGlnLysThrProGlyArgIleArgSerSer 240
QY 694 GGACACTTATCTACAAGCTCTTTTACA----- 720
Db 241 GlyValLeuSerMetValPheTyrThrAspSerAlaIleAlaLysGluGlyPheSerAla 260
QY 721 -----TTTAAGACAAAGTGGATGTTATGGA 744
Db 261 AsnTyrSerValLeuGlnSerSerIleSerGluAspPheLys-----CysMetGlu 277
QY 745 ACACCTGGGATGGAGTCTGGTGTGATCGCGGATCCCTCAAAATACACAGCATCATCTGTGCTG 804
Db 278 AlaLeuGlyMetGluSerGlyGluIleHisSerAspGlnIleThrAlaSer----- 295
QY 805 GAGTGGACTGACCCACACAGGCGCAAGAACACTGTTGAAACCCAAAAGCCAGCTGAAA 864
Db 296 -----GlnTyrGlyThrAsnTrpSerValGluArgSerArgLeuAsn 309
QY 865 AAACCTGGACCCCTTGGGCTCTTTGCCACTGATGAATACACAGTGGTACAAATAGAT 924
Db 310 TyrProGluAsnGlyTrpThrProGlyGluAspSerTyrLysGluTrpIleGlnValAsp 329
QY 925 TTGAATAAGGAAAGAAAATAACAGGCATTTATAACACTGGA-----TCCACCATGCTG 978
Db 330 LeuGlyLeuLeuArgPheValThrAlaValGlyThrGlnGlyAlaIleSerLysGluThr 349
QY 979 GAGCAACAATTACTATGTGTCTGCCCTACAGAATCCCTGTGATGATGATGGGAGAAATGG 1038
Db 979 GAGCAACAATTACTATGTGTCTGCCCTACAGAATCCCTGTGATGATGATGGGAGAAATGG 1038

Db 333 LeuArgPheValThrAlaValGlyThrGlnGlyAlaIleSerLysGluThrLysLysLys 352
QY 988 TACTATGTCTCCCTACAACTCTGACAGTATGATGGCAGAAATGGACGTGTGAC 1047
Db 353 TyrTyrValLysThrTyrArgValAspIleSerSerAsnGlyGluAspTTPilleThrLeu 372
QY 1048 AGAGAGCTGTGTGGAGCAGATAGATATTTCAAGGAACAAAGATATATCACCAGGAT 1107
Db 373 LysGluGly-----AsnLysAlaIlePheGlnGlyAsnThrAsnProThrAspVal 390
QY 1108 GTGGTAAATACCTTTGGCCACCAATATTGACAGTATTTATTAGAGTAATCCACCCAA 1167
Db 391 ValPheGlyValPheProLysProLeuIleThrArgPheValArgIleLysProAlaSer 410
QY 1168 TGGCAGCAGAAATGTCATCAAAATGGAGCTGCTCGATGCTCAGTTATTCTTAAAGGT 1227
Db 411 TTPGluThrGlyIleSerMetArgPheGluValTyrGlyCys----- 424
QY 1228 CGTCTCCAAACTTACTCAACCTCCA 1254
Db 425 -----LysIleThrAspTyrPro 430
RESULT 13
US-08-746-111-5
; Sequence 5, Application US/08746111
; Patent No. 6066778
; GENERAL INFORMATION:
; APPLICANT: Ginsburg, David
; APPLICANT: Cui, Jisong
; TITLE OF INVENTION: Compositions And Methods For Screening
; TITLE OF INVENTION: Compounds For Anticoagulant Activity
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,111
; FILING DATE: 06-NOV-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-746-111-5
Alignment Scores:
Pred. No.: 6.35e-25 Length: 2183
Score: 355.50 Matches: 79
Percent Similarity: 54.31% Conservative: 28
Best Local Similarity: 40.10% Mismatches: 75
Query Match: 8.96% Indels: 15
DB: 3 Gaps: 4
US-10-060-830-2 (1-2190) x US-08-746-111-5 (1-2183)

QY 631 GTAATTAGTAAGATATTCCTTATTAAGAAAGTCTTTTGGCTAACACGTCACATCTGTG 690
Db 1995 IleValAlaArgTyrIleArgIleHisProThrLysSerTyrAsnArgProThr----- 2012
QY 691 GTGGACACATATCTACAGTCTTTTACATTTTAAGACAACTGGATCTTATGGACACTG 750
Db 2013 -----LeuArgLeuGluLeuGlnGlyCysGluValAsnGlyCysSerThrProLeu 2029
QY 751 GGGATGGAGTCTGTGTGATCGCGATCCTCAATAAATACAGCATCTGTCTCTGGAG--- 807
Db 2030 GlyLeuGluAspGlyArgIleGlnAspLysGlnIleThrAlaSerSerPheLysLysSer 2049
QY 808 ---TGGACTGACACACAGGCGCAAGACAGTGGAAACCCCAAGCCAGCGCTGAAA 864
Db 2050 TTPTrpGlyAspTyr-----TTPGluProSerLeuAlaArgLeuAsn 2063
QY 865 AAACCTGGA-----CCGCTTGGCTGCTTTTGGCTGATGATGATACCATCTGTACAA 918
Db 2064 AlaGlnGlyArgValAsnAlaTTPGlnAlaLysAlaAsnAsnLysGlnTTPLeuGln 2083
QY 919 ATAGATTGTAATAAGAAAGAAATACAGGCATTATAACCACTGGATCCACCATGGTG 978
Db 2084 ValAspLeuLeuLysIleLysLysValThrAlaIleValThrGlnGlyCysLysSerLeu 2103
QY 979 GAGCACATTTACTGTCTGCTCTACAGAAATCCTGTACAGTATGATGGCAGAAATGG 1038
Db 2104 SerSerGluMetTyrValLysSerTyrSerIleGlnTyrSerAspGlnGlyValAlaTTP 2123
QY 1039 ACTGTGTACAGAGCCCTGGTGGAGCAAGATAGATATTTCAGGAAACAAAGATTAT 1098
Db 2124 LysProTyrArgGlnLysSerSerMetValAspLysIlePheGluGlyAsnSerAsnThr 2143
QY 1099 CACCAGGATGTGCGTAATACTTTTGGCCACCAATATTGACAGCTTTTATAGAGTGAAT 1158
Db 2144 LysGlyHisMetLysAsnPhePheAsnProProllelleSerArgPheIleArgIleIle 2163
QY 1159 CTACCCCAATGGCAGCAGAAATTCGCATGAAATGGAGCTGCTCGGATGT 1209
Db 2164 ProlysThrTTPAsnGlnSerIleAlaLeuArgLeuGluPheGlyCys 2180
RESULT 14
US-08-480-229C-10
; Sequence 10, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quetermous, Thomas
; APPLICANT: Hogan, Bridgid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0026-999
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-229C-10

Alignment Scores:
Pred. No.: 5,66e-25 Length: 480
Score: 352.00 Matches: 75
Percent Similarity: 59.30% Conservative: 27
Best Local Similarity: 43.60% Mismatches: 66
Query Match: 8.87% Indels: 4
DB: 2 Gaps: 2

US-10-060-830-2 (1-2190) x US-08-480-229C-10 (1-480)

QY 700 TTATCTACAAGCTCTTTTACATTTTAAAGACAAGTGGATGTTTGAACACTGGGATGGAG 759
Db 307 LeuArgMetGluLeuLeuGlyCysGlnLeuSerGlyCysSerGluProLeuGlyMetLys 326
QY 760 TCTGTGTGTCGCGGATCCTCAATAACAGCATCATCTGTCTGGAGTGGAGTACACCAC 819
Db 327 SerGlyHisIleGlnAspTyrGlnIleThrAlaSerSerValPheArgThrLeuAsnMet 346
QY 820 ACAGGGCAAGAGACAGTGTGGAACCCAAAGCCAGCGTGAACAAACCTGGACCG--- 876
Db 347 -----AspMetPheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysVal 364
QY 877 ---CCTTGGCTGCTTTTCCCACTCATCAATACAGTGTGTTACAAATAGATTGAATAAG 933
Db 365 AsnAlaTrpThrSerGlyHisAsnAspGlnSerGlnIleThrLeuGlnValAspLeuVal 384
QY 934 GAAAGAAATAACAGCATTAACCACTGGATCCACCATGGTGGAGCACAATTAATCTAT 993
Db 385 ProThrLysValThrGlyIleThrGlnGlyAlaLysAspPheGlyHisValGlnPhe 404
QY 994 GTGTCTGCTACAGATCCTGTACAGTGTATGGGAGAAATGAGCTGTGTACAGAGAG 1053
Db 405 ValGlySerTyrLysLeuAlaTyrSerAsnAspGlyGluHisTrpMetValHisGlnAsp 424
QY 1054 CTTGTGTGGAGCAAGATATTTCAAGAAACAAAGATTATCACCAGGATGTCGT 1113
Db 425 GluLysGlnArgLysAspLysValPheGlnGlyAsnPheAspAsnAspThrHisArgLys 444
QY 1174 CAGAAATGGCATGAAATGAGCTGTCGGATGT 1209
Db 465 GlyArgIleThrLeuArgSerGluLeuLeuGlyCys 476

RESULT 15

US-08-659-235C-10
Sequence 10, Application US/08659235C
Patent No. 5877281
GENERAL INFORMATION:
APPLICANT: Quettermous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York

STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/659,235C
FILING DATE: 05-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0034-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-659-235C-10

Alignment Scores:
Pred. No.: 6,66e-25 Length: 480
Score: 352.00 Matches: 75
Percent Similarity: 59.30% Conservative: 27
Best Local Similarity: 43.60% Mismatches: 66
Query Match: 8.87% Indels: 4
DB: 2 Gaps: 2

US-10-060-830-2 (1-2190) x US-08-659-235C-10 (1-480)

QY 700 TTATCTACAAGCTCTTTTACATTTTAAAGACAAGTGGATGTTTGAACACTGGGATGGAG 759
Db 307 LeuArgMetGluLeuLeuGlyCysGlnLeuSerGlyCysSerGluProLeuGlyMetLys 326
QY 760 TCTGTGTGTCGCGGATCCTCAATAACAGCATCATCTGTCTGGAGTGGAGTACACCAC 819
Db 327 SerGlyHisIleGlnAspTyrGlnIleThrAlaSerSerValPheArgThrLeuAsnMet 346
QY 820 ACAGGGCAAGAGACAGTGTGGAACCCAAAGCCAGCGTGAACAAACCTGGACCG--- 876
Db 347 -----AspMetPheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysVal 364
QY 877 ---CCTTGGCTGCTTTTCCCACTCATCAATACAGTGTGTTACAAATAGATTGAATAAG 933
Db 365 AsnAlaTrpThrSerGlyHisAsnAspGlnSerGlnIleThrLeuGlnValAspLeuVal 384
QY 934 GAAAGAAATAACAGCATTAACCACTGGATCCACCATGGTGGAGCACAATTAATCTAT 993
Db 385 ProThrLysValThrGlyIleThrGlnGlyAlaLysAspPheGlyHisValGlnPhe 404
QY 994 GTGTCTGCTACAGATCCTGTACAGTGTATGGGAGAAATGAGCTGTGTACAGAGAG 1053
Db 405 ValGlySerTyrLysLeuAlaTyrSerAsnAspGlyGluHisTrpMetValHisGlnAsp 424
QY 1054 CTTGTGTGGAGCAAGATATTTCAAGAAACAAAGATTATCACCAGGATGTCGT 1113
Db 425 GluLysGlnArgLysAspLysValPheGlnGlyAsnPheAspAsnAspThrHisArgLys 444
QY 1114 AATAACTTTTGGCCACCAATTTATTCAGCTTTTATTAGAGTGAATCCCAATGGCAG 1173
Db 445 AsnValIleAspProIleTyrAlaArgPheIleArgIleLeuProIleTyrSerTrpTyr 464
QY 1174 CAGAAATGGCATGAAATGAGCTGTCGGATGT 1209
Db 465 GlyArgIleThrLeuArgSerGluLeuLeuGlyCys 476

```
FT VARSPLIC 719 1156
FT CONFLICT 351 351
FT CONFLICT 680 682
FT CONFLICT 689 689
FT CONFLICT 725 727
FT CONFLICT 746 746
FT CONFLICT 912 912
FT CONFLICT 1089 1089
SQ SEQUENCE 1156 AA; 127938 MW; 5AA90E8938A84A0A CRC64;

A (IN SHORT ISOFORM).
MISSING (IN SHORT ISOFORM).
A -> V (IN REF. 1).
ACM -> GCL (IN REF. 1).
H -> Q (IN REF. 1 AND 2).
MAQ -> LAH (IN REF. 1).
L -> F (IN REF. 1).
N -> D (IN REF. 1).
G -> C (IN REF. 2).

Alignment Scores:
Pred. No.: 9.62 Length: 1156
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-1115 (1-60) x HTS_DROME (1-1156)
QY 23 GCTCCAGCGTCTCGAGCAGC 3
DB 1062 AlaProAlaSerSerSer 1068
|||||
```

Search completed: January 21, 2003, 14:49:59
Job time : 3.84252 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:43:31 ; Search time 6.18898 Seconds

(without alignments)

3995.112 Million cell updates/sec

Title: US-10-060-830-1115

Perfect score: 20

Sequence: 1 ctgctgctgagacgtgg.....gtggacacactgtactagc 60

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1343043

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+azp.model -DEV=xlh
-O=/cgn2_1/USPRO_spool/US10060830/runat_16012003_092702_1498/app_query.fasta_1.1109
-DB=SPREMBL_21 -Qfmt=fastan -SUFFIX=olin2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn_1_1_57_@runat_16012003_092702_1498 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

SPREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	775	4	Q96pd2 homo sapien

2	20	100.0	775	4	Q8TDX2
3	12	60.0	769	11	Q91ZV3
4	10	50.0	34	4	Q96IX0
5	10	50.0	769	11	Q91ZV2
6	9	45.0	364	4	Q14089
7	8	42.1	2237	9	Q8SCY1
8	7	35.0	63	16	Q8YQP2
9	7	36.8	67	8	Q36734
10	7	36.8	96	5	Q95RZ7
11	7	36.8	102	11	Q9D7Q2
12	7	35.0	107	5	Q45433
13	7	35.0	107	5	Q94054
14	7	36.8	110	10	Q9FDY5
15	7	35.0	155	17	Q8ZX18
16	7	35.0	227	16	Q9RWX2
17	7	36.8	249	10	Q9SCM4
18	7	36.8	261	10	Q9XG96
19	7	36.8	297	5	Q9V5R4
20	7	36.8	307	4	Q96DB3
21	7	36.8	323	10	Q43529
22	7	36.8	347	8	Q9B1Q7
23	7	36.8	347	8	Q9B0T5
24	7	36.8	347	8	Q78760
25	7	36.8	347	8	Q9T9W2
26	7	36.8	347	8	Q9T9Y9
27	7	36.8	347	8	Q9T9X1
28	7	36.8	347	8	Q9B2Z8
29	7	36.8	347	8	Q9B2Z6
30	7	36.8	347	8	Q9B2Y8
31	7	36.8	347	8	Q9B2Y6
32	7	36.8	347	8	Q9B2W4
33	7	36.8	347	8	Q94PV3
34	7	36.8	347	8	Q9B2Z3
35	7	36.8	347	8	Q9B186
36	7	36.8	347	8	Q9B2X3
37	7	36.8	347	8	Q8WCY6
38	7	36.8	347	8	Q8WCX4
39	7	36.8	347	8	Q8WCW8
40	7	36.8	347	8	Q79988
41	7	36.8	347	8	Q9T3Q5
42	7	36.8	347	8	Q78759
43	7	36.8	347	8	Q36733
44	7	36.8	360	17	Q9HJ65
45	7	35.0	391	13	P87380

ALIGNMENTS

RESULT 1
ID Q96PD2 PRELIMINARY; PRT; 775 AA.
AC Q96PD2;
DT 01-DEC-2001 (TREMREL. 19, Created)
DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
GN ESDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury";
RL J. Biol. Chem. 276:34105-34114(2001).
CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.

Db 7 LeuLeuLeuGluAspAlaGlyAlaGlnGln 16

RESULT 5

Q91ZV2 PRELIMINARY; PRT; 769 AA.

AC Q91ZV2

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Endothelial and smooth muscle cell-derived neuropilin-like protein.

GN ESDN.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY;

RX PubMed=11447234;

RA Kobuke K., Fukawa Y., Sugai M., Tanigaki K., Ohashi N.,

RA Matsumori A., Sasayana S., Honjo T., Tashiro K.;

RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular

RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is

RT Up-regulated after Vascular Injury.;"

RT J. Biol. Chem. 276:34105-34114(2001).

CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL

CC REGULATORS.

CC EMBL: AF387549; AAL30180.1; -

DR InterPro: IPR000859; CUB_domain.

DR InterPro: IPR000421; FA58_C.

DR InterPro: IPR004043; LCCL_dom.

DR Pfam: PF00431; CUB; 1.

DR Pfam: PF00754; F5_F8_type_C; 1.

DR PROSITE: PS01180; CUB; 1.

DR DNA-binding; transcription regulation.

KW

SQ SEQUENCE 769 AA; 83915 MW; 5B7936C8FA063A93 CRC64;

Alignment Scores:

Pred. No.:	0.0276	Length:	769
Score:	100.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	50.00%	Indels:	0
DB:	11	Gaps:	0

US-10-060-830-1115 (1-60) x Q91ZV2 (1-769)

QY 31 GGTGGTGGTGGACACACTGTACTAGGC 60

|||||

Db 66 GlyAspGlyCysGlyHisThrValLeuGly 75

RESULT 6

Q14089 PRELIMINARY; PRT; 364 AA.

AC Q14089

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Hypothetical 40.0 kDa protein (fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Shibata T.;

RT "unpublished.;"

RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.

DR EMBL: D29810; BAA18909.1; -

DR InterPro: IPR000859; CUB_domain.

DR InterPro: IPR000421; FA58_C.

DR InterPro: IPR004043; LCCL_dom.

DR Pfam: PF00431; CUB; 1.

DR Pfam: PF00754; F5_F8_type_C; 1.

DR SMART: SM00042; CUB; 1.

DR PROSITE: PS01180; CUB; 1.

KW Hypothetical protein.

FT NON_TER 1

SQ SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;

Alignment Scores:

Pred. No.:	0.349	Length:	364
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.00%	Indels:	0
DB:	4	Gaps:	0

US-10-060-830-1115 (1-60) x Q14089 (1-364)

QY 34 GATGGATGGACACACTGTACTAGGC 60

|||||

Db 1 AspGlyCysGlyHisThrValLeuGly 9

RESULT 7

Q8SCY1 PRELIMINARY; PRT; 2237 AA.

ID Q8SCY1

AC Q8SCY1

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE PHK2181.

OS Pseudomonas phage phiK2.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.

OX NCBI_TaxID=169683;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21914557; PubMed=11916376;

RA Mesyanzhinov V.V., Robben J., Grymonprez B., Kostyuchenko V.A.,

RA Bourkaltseva M.V., Sykilla N.N., Krylov V.N., Volckaert G.;

RT "The genome of bacteriophage phiK2 of Pseudomonas aeruginosa.;"

RL J. Mol. Biol. 317:1-19(2002).

RN [2]

RP SEQUENCE FROM N.A.

RA Mesyanzhinov V.V., Robben J., Grymonprez B., Kostyuchenko V.A.,

RA Bourkaltseva M.V., Sykilla N.N., Krylov V.V., Volckaert G.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF399011; AAL83082.1; -

SQ SEQUENCE 2237 AA; 245756 MW; 6CA92E901B372F04 CRC64;

Alignment Scores:

Pred. No.:	3.11	Length:	2237
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	42.11%	Indels:	0
DB:	9	Gaps:	0

US-10-060-830-1115 (1-60) x Q8SCY1 (1-2237)

QY 51 AGTGTGTCACATCCATCCTTG 28

|||||

Db 1695 SerValSerThrSerIleThrLeu 1702

RESULT 8

Q8YQP2 PRELIMINARY; PRT; 63 AA.

ID Q8YQP2

AC Q8YQP2

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Hypothetical protein Asl3778.

GN Asl3778.

OS Anabaena sp. (strain PCC 7120).

OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyano bacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 DR EMBL: AP003594; BAB75477.1; --
 KW Hypochemical protein; Complete proteome.
 SQ SEQUENCE 63 AA; 7285 MW; F31AC141EA5444FD CRC64;

Alignment Scores:
 Pred. No.: 57.8 Length: 63
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 35.00% Indels: 0
 DB: 16 Gaps: 0

US-10-060-830-1115 (1-60) x Q8YQP2 (1-63)
 QY 17 CTGAGCCCAAGGATGATG 37
 Db 8 LeuGluProSerLysValMet 14
 |||||
 |||||

RESULT 9
 Q36734
 ID Q36734 PRELIMINARY; PRT; 67 AA.
 AC Q36734;
 DT 01-NOV-1996 (TREMREL. 01, Created)
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE NADH dehydrogenase subunit 2 (Fragment).
 GN CLONE 39-1.
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92118019; PubMed=1370613;
 RA Lin F.H., Lin R., Wisniewski H.M., Hwang Y.W., Grundke-Iqbal I.,
 RA Healy-Louie G., Iqbal K.;
 RT "Detection of point mutations in codon 331 of mitochondrial NADH
 dehydrogenase subunit 2 in Alzheimer's brains.";
 RL Biochem. Biophys. Res. Commun. 182:238-246(1992).
 DR EMBL: S75896; AAB20906.1; --
 KW Mitochondrion.
 FT NON_TER 1
 SQ SEQUENCE 67 AA; 7928 MW; AF8C8BA74943EFF1 CRC64;

Alignment Scores:
 Pred. No.: 57.3 Length: 67
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 8 Gaps: 0

US-10-060-830-1115 (1-60) x Q36734 (1-67)
 QY 45 TCCACATCCATCACTTCG 25
 Db 24 SerThrSerIleThrLeuLeu 30
 |||||
 |||||

RESULT 10

Q95RZ7
 ID Q95RZ7 PRELIMINARY; PRT; 96 AA.
 AC Q95RZ7;
 DT 01-DEC-2001 (TREMREL. 19, Created)
 DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE HL04994p.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
 RL Submitted (OCI-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY061024; AAL28572.1; --
 SQ SEQUENCE 96 AA; 10179 MW; 4DE564B0E38CF831 CRC64;

Alignment Scores:
 Pred. No.: 54.6 Length: 96
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 36.84% Indels: 0
 DB: 5 Gaps: 0

US-10-060-830-1115 (1-60) x Q95RZ7 (1-96)
 QY 56 AGTAGTGTCACATCCA 36
 |||||
 |||||

Db 66 SerThrValCysProHisPro 72
 |||||

RESULT 11
 Q9D7Q2
 ID Q9D7Q2 PRELIMINARY; PRT; 102 AA.
 AC Q9D7Q2;
 DT 01-JUN-2001 (TREMREL. 17, Created)
 DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
 DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
 DE 2300002G24RIK protein.
 GN 2300002G24RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
 RA Schrim L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK009010; BAB26023.1; -.
DR MGD; MGI:1921425; 2300002G24Rik.
SQ SEQUENCE 102 AA; 9749 MW; 8BFE5A8D6AE78857 CRC64;

Alignment Scores:
Pred. No.: 54.1 Length: 102
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 11 Gaps: 0

US-10-060-830-1115 (1-60) x Q9D7Q2 (1-102)

QY 23 GCTCCAGCGTCCTCGAGCAGC 3
Db 26 AlaProAlaSerSerSerSer 32

RESULT 12
O45433 PRELIMINARY; PRT; 107 AA.
ID O45433
AC O45433;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F32B6.7 protein.
GN F32B6.7
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Basham V.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z81074; CAB03041.1; -.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
SQ SEQUENCE 107 AA; 11759 MW; 9E02E9CB205CD2B5 CRC64;

Alignment Scores:
Pred. No.: 53.8 Length: 107
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 5 Gaps: 0

US-10-060-830-1115 (1-60) x O45433 (1-107)

QY 16 GCTGGAGCCCGCAGCGTGTAT 36
Db 93 AlaGlyAlaGlnGlnGlyAsp 99

RESULT 13
O94054 PRELIMINARY; PRT; 107 AA.
ID O94054
AC O94054;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T13P2.12 protein.
GN T13P2.12
OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Swinburne J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z81122; CAB03363.1; -.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
SQ SEQUENCE 107 AA; 11746 MW; 8B2E9CB205CD2B5 CRC64;

Alignment Scores:
Pred. No.: 53.8 Length: 107
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.00% Indels: 0
DB: 5 Gaps: 0

US-10-060-830-1115 (1-60) x Q94054 (1-107)

QY 16 GCTGGAGCCCGCAGCGTGTAT 36
Db 93 AlaGlyAlaGlnGlnGlyAsp 99

RESULT 14
Q9FDY5 PRELIMINARY; PRT; 110 AA.
ID Q9FDY5
AC Q9FDY5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Photosystem I subunit PSI-E-like protein (F16A16.140).
GN F16A16.140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu S.X., Sakano H., Pham P.K., Yamada K., Banh J., Etgu P., Lee J.M.,
RA Toriumi M., Yu G., Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene F16A16.140; AT4g28750 (GI:7269730).";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Chao Q., Chen H., Kim C., Ecker J.;
RT "Full length cDNA sequences of Arabidopsis thaliana.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF326861; AAG41443.1; -.
DR EMBL; AF324694; AAG40045.1; -.
DR HSSP; Q9WNP1; IQP2.
DR InterPro; IPR003375; PSI_Psae.
DR Pfam; PF02427; PSI_Psae; 1.
SQ SEQUENCE 110 AA; 11710 MW; EAF5F498E2FE933 CRC64;

Alignment Scores:
Pred. No.: 53.6 Length: 110
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-1115 (1-60) x Q9FDY5 (1-110)

QY 23 GCTCCAGCGTCTCGAGCAGC 3
 |||||
 Db 17 AlaProAlaSerSerSer 23

RESULT 15

Q8ZX18
 ID Q8ZX18 PRELIMINARY; PRT; 155 AA.
 AC Q8ZX18;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE PAREP8
 GN PAE1515.
 OS Pyrobaculum aerophilum.
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
 OC Thermoproteaceae; Pyrobaculum.
 OX NCBI_TaxID=13773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
 RX PubMed=11792869;
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
 RA Miller J.H.;
 RT "genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
 RT aerophilum.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
 DR EMBL; AE009825; AAL63531.1; -;
 KW Complete proteome.
 SQ SEQUENCE 155 AA; 17000 MW; EF1281E053AB476D CRC64;

Alignment Scores:
 Pred. No.: 51.1 Length: 155
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 35.00% Indels: 0
 DB: 17 Gaps: 0

US-10-060-830-1115 (1-60) x Q8ZX18 (1-155)

QY 3 GCTCGAGGAGCGTGGAGC 23
 |||||
 Db 89 AlaAlaArgGlyArgTrpSer 95

Search completed: January 21, 2003, 14:52:26
 Job time : 8.18898 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:46:52 ; Search time 2.17323 Seconds
(without alignments)
1624.657 Million cell updates/sec

Title: us-10-060-830-1115
Perfect score: 20
Sequence: 1 ctgctgctgagacgctgg.....gtggacacactgtactagc 60

Scoring table:

OLIGO	Xgapop 60.0	Xgapext 60.0
Ygapop 60.0	Ygapext 60.0	
Fgapop 6.0	Fgapext 7.0	
Delopt 6.0	Delext 7.0	

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438675

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-O/cgn2_1/USPTO.spool/US10060830/runat_16012003_092703_1523/app_query.fasta_1.1109
-DB-Issued Patents_AA -QFMT-fastan -SUFFIX-olin2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-oligo -TRANS-human40.cd1
-LIST=45 -DOCALIGN=200 -THR_SCORE-quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830@cgn2_1.8@runat_16012003_092703_1523 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	7	36.8	347	4	US-09-097-889-14
C 2	7	36.8	422	4	US-09-071-035-484
C 3	7	36.8	449	4	US-09-071-035-482
C 4	6	31.6	17	3	US-08-974-899-23
C 5	6	31.6	38	4	US-08-469-260A-331
C 6	6	31.6	42	2	US-08-476-866-23
C 7	6	30.0	52	4	US-09-346-860-3
C 8	6	31.6	60	4	US-09-149-476-678
C 9	6	31.6	60	4	US-09-149-476-682
C 10	6	30.0	82	4	US-09-382-155-10
C 11	6	30.0	82	4	US-09-074-044A-10
C 12	6	31.6	121	3	US-08-974-899-24

C 13	31.6	158	3	US-09-053-197A-23	Sequence 23, Appl
C 14	31.6	158	4	US-09-085-761A-23	Sequence 23, Appl
C 15	31.6	159	2	US-08-606-143-45	Sequence 45, Appl
C 16	31.6	182	2	US-08-606-143-44	Sequence 44, Appl
C 17	31.6	203	4	US-09-149-476-383	Sequence 383, Appl
C 18	31.6	210	4	US-09-162-184-33	Sequence 33, Appl
C 19	31.6	210	4	US-09-489-777A-33	Sequence 33, Appl
C 20	30.0	248	2	US-08-755-559-1	Sequence 1, Appl
C 21	30.0	248	3	US-09-210-474-1	Sequence 1, Appl
C 22	30.0	248	4	US-08-539-774-1	Sequence 1, Appl
C 23	30.0	260	4	US-09-134-001C-4507	Sequence 4507, Ap
C 24	30.0	307	4	US-09-105-390-42	Sequence 42, Appl
C 25	31.6	326	1	US-07-603-133B-22	Sequence 22, Appl
C 26	31.6	326	1	US-07-603-133B-23	Sequence 23, Appl
C 27	31.6	326	3	US-08-089-397A-13	Sequence 13, Appl
C 28	31.6	326	3	US-08-089-397A-14	Sequence 14, Appl
C 29	31.6	329	2	US-08-913-477-2	Sequence 2, Appl
C 30	31.6	329	2	US-08-913-477-4	Sequence 4, Appl
C 31	30.0	332	4	US-09-105-390-58	Sequence 58, Appl
C 32	30.0	371	4	US-09-382-155-22	Sequence 22, Appl
C 33	30.0	371	4	US-09-074-044A-22	Sequence 22, Appl
C 34	31.6	388	4	US-09-188-930-275	Sequence 275, App
C 35	31.6	406	4	US-09-719-083A-4	Sequence 4, Appl
C 36	31.6	410	1	US-07-945-283-4	Sequence 4, Appl
C 37	31.6	454	4	US-09-134-001C-3547	Sequence 3547, Ap
C 38	31.6	459	1	US-08-220-151-12	Sequence 12, Appl
C 39	31.6	459	1	US-08-220-151-14	Sequence 14, Appl
C 40	31.6	459	1	US-08-413-118-12	Sequence 12, Appl
C 41	31.6	459	1	US-08-413-118-14	Sequence 14, Appl
C 42	31.6	459	3	US-08-473-446-12	Sequence 12, Appl
C 43	31.6	459	3	US-08-473-446-14	Sequence 14, Appl
C 44	31.6	459	4	US-09-213-053-6	Sequence 6, Appl
C 45	31.6	480	2	US-08-913-477-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1

US-09-097-889-14 Application US/09097889
; Sequence 14, Application US/09097889
; Patent No. 6218117

GENERAL INFORMATION:

APPLICANT: Heirnstadt, Corrina
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Davis, Robert E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AGENTS THAT QUANTITATIVELY ALTER DETECTABLE EXTRAMITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS
TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

Result

No.	Score	Query Match	Length	ID	Description
C 1	7	36.8	347	4	US-09-097-889-14
C 2	7	36.8	422	4	US-09-071-035-484
C 3	7	36.8	449	4	US-09-071-035-482
C 4	6	31.6	17	3	US-08-974-899-23
C 5	6	31.6	38	4	US-08-469-260A-331
C 6	6	31.6	42	2	US-08-476-866-23
C 7	6	30.0	52	4	US-09-346-860-3
C 8	6	31.6	60	4	US-09-149-476-678
C 9	6	31.6	60	4	US-09-149-476-682
C 10	6	30.0	82	4	US-09-382-155-10
C 11	6	30.0	82	4	US-09-074-044A-10
C 12	6	31.6	121	3	US-08-974-899-24

INFORMATION FOR SEQ ID NO: 14:

STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,899
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031971
FILING DATE: 11/27/96
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 40,378
REFERENCE/DOCKET NUMBER: P1014R1
TELEPHONE: 650/225-1994
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-974-899-23

Alignment Scores:
Pred. No.: 146 Length: 17
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.58% Indels: 0
DB: 3 Gaps: 0

US-10-060-830-1115 (1-60) x US-08-974-899-23 (1-17)

QY 23 GCTCCAGCGTCCTCGAGC 6
Patent No. 6451578
Db 3 AlaProAlaSerSerSer 8

RESULT 5

US-08-469-260A-331
Sequence 331, Application US/08469260A

Patent No. 6451578

GENERAL INFORMATION:

APPLICANT: JOHN N. SIMONS
APPLICANT: TAMI J. PILOT-MATIAS
APPLICANT: GEORGE J. DAWSON
APPLICANT: GEORGE G. SCHLAUDER
APPLICANT: SURESH M. DESAI
APPLICANT: THOMAS P. LEARY
APPLICANT: ANTHONY SCOTT MUEHROFF
APPLICANT: JAMES C. ERKER
APPLICANT: SHERI L. BUIJK
APPLICANT: ISA K. MUSHAWAR
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
NUMBER OF SEQUENCES: 715
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: 100 ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,260A
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/424,550
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: POREMSKI, PRISCILLA E.

REGISTRATION NUMBER: 33,207

REFERENCE/DOCKET NUMBER: 5527.PC.01

TELEPHONE: 708-937-6365

TELEFAX: 708-938-2623

INFORMATION FOR SEQ ID NO: 331:

SEQUENCE CHARACTERISTICS:

LENGTH: 38 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-469-260A-331

Alignment Scores:

Pred. No.: 131 Length: 38
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.58% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-1115 (1-60) x US-08-469-260A-331 (1-38)

QY 45 TCCACATCCATCACCCTTG 28

Db 17 SerThrSerIleThrLeu 22

RESULT 6

US-08-476-866-23

Sequence 23, Application US/08476866

Patent No. 5994339

GENERAL INFORMATION:

APPLICANT: CRAPO, JAMES D.

APPLICANT: FRIDOVICH, IRWIN

APPLICANT: OURY, TIM

APPLICANT: DAY, BRIAN J.

APPLICANT: FOLZ, RODNEY J.

APPLICANT: FREEMAN, BRUCE A.

TITLE OF INVENTION: SUPEROXIDE DISMUTASE AND MIMETICS THEREOF

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHVE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/476,866

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/322,766

FILING DATE: 13-OCT-1994

APPLICATION NUMBER: US 08/136,207

FILING DATE: 15-OCT-1993

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
 NAME: WILSON, MARY J.
 REGISTRATION NUMBER: 32,955
 REFERENCE/DOCKET NUMBER: 1579-74
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 816-4000
 TELEFAX: (703) 816-4100
 TELEX: 200797 NIXN UR
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 42 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-476-866-23

Alignment Scores:
 Pred. No.: 129 Length: 42
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.58% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x US-08-476-866-23 (1-42)

Qy 20 CCAGGCTCTCGAGCAGC 3
 |||||
 Db 1 ProLaserSerSerSer 6
 RESULT 7
 US-09-346-860-3
 Sequence 3, Application US/09346860
 Patent No. 6221641
 GENERAL INFORMATION:
 APPLICANT: Khosla, Chaitan
 APPLICANT: Lau, Janice
 APPLICANT: Pohl, Nicola L.
 TITLE OF INVENTION: METHOD FOR MAKING POLYKETIDES
 FILE REFERENCE: 30062-20025.00
 CURRENT APPLICATION NUMBER: US/09/346,860
 CURRENT FILING DATE: 1999-07-02
 NUMBER OF SEQ ID NOS: 15
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 3
 LENGTH: 52
 TYPE: PRT
 ORGANISM: Streptomyces sp.
 US-09-346-860-3

Alignment Scores:
 Pred. No.: 125 Length: 52
 Score: 6.00 Matches: 6
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 30.00% Indels: 0
 DB: 4 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-346-860-3 (1-52)

Qy 7 CTCGAGGACGTGGAGCC 24
 |||||
 Db 18 LeuGluAspAlaGlyAla 23
 RESULT 8
 US-09-149-476-678
 Sequence 678, Application US/09149476
 Patent No. 6420526
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: 186 Human Secreted proteins
 FILE REFERENCE: PZ002P1
 CURRENT APPLICATION NUMBER: US/09/149,476

CURRENT FILING DATE: 1998-09-08
 EARLIER APPLICATION NUMBER: PCT/US98/04493
 EARLIER FILING DATE: 1998-03-06
 EARLIER APPLICATION NUMBER: 60/040,162
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,333
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/038,621
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,626
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,334
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,336
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,163
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/047,600
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,615
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,597
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,502
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,633
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,583
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,617
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,618
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,503
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,592
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,581
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,584
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,500
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,587
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,492
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,598
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,613
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,582
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,596
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,612
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,632
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,601
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/043,580
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,568
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,314
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,569
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,311
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,671
 EARLIER FILING DATE: 1997-04-11

[illegible]

EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

Alignment Scores:
Pred. No.: 123 Length: 60
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.58% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-149-476-682 (1-60)

QY 50 GTGTGTCACATCCATCA 33
Db 35 ValcysProHisProser 40

RESULT 10

US-09-382-155-10
Sequence 10, Application US/09382155B
Patent No. 6160095
GENERAL INFORMATION:
APPLICANT: CHAUDHARY, PREET M
APPLICANT: HOOD, LEROY
TITLE OF INVENTION: PROTEINS CAPABLE OF REGULATING NF-KB, JNK AND APOPTOSIS
TITLE OF INVENTION: PATHWAYS AND METHODS OF USING THE SAME
FILE REFERENCE: Chaudhary
CURRENT FILING DATE: 1999-08-24
EARLIER APPLICATION NUMBER: 09/074,044

EARLIER FILING DATE: 1998-05-07
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 82
TYPE: PRT
ORGANISM: 'Axial Seamount' polynoid polychaete
US-09-382-155-10

Alignment Scores:
Pred. No.: 118 Length: 82
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 30.00% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-382-155-10 (1-82)

QY 7 CTCGAGGACGCTGGAGCC 24

Db 47 LeuGluaspAlaGlyAla 52

RESULT 11

US-09-074-044A-10
Sequence 10, Application US/09074044A
Patent No. 6207458

GENERAL INFORMATION:
APPLICANT: CHAUDHARY, PREET M
APPLICANT: HOOD, LEROY
TITLE OF INVENTION: PROTEINS CAPABLE OF REGULATING NF-KB, JNK AND APOPTOSIS
TITLE OF INVENTION: APOPTOSIS PATHWAYS AND METHODS OF USING THE SAME
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 2405 GRAND BLVD., SUITE 400
CITY: KANSAS CITY
STATE: MISSOURI
COUNTRY: USA
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,044A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: COLLINS, JOHN M
REGISTRATION NUMBER: 26,262
REFERENCE/DOCKET NUMBER: 26588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 816/474-9050
TELEFAX: 816/474-9057
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 82 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: MOLLUSCUM CONTAGIOSUM VIRUS
US-09-074-044A-10

Alignment Scores:
Pred. No.: 118 Length: 82
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 30.00% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-074-044A-10 (1-82)

Qy 7 CTCGAGGAGCTGGAGCC 24
Db 47 LeuGluspAlaGlyAla 52

RESULT 12
US-08-974-899-24
; Sequence 24, Application US/08974899
; Patent No. 6037454
; GENERAL INFORMATION:
; APPLICANT: Presta, Leonard G.
; APPLICANT: Jardieu, Paula M.
; TITLE OF INVENTION: Humanized Anti-CD11a Antibodies
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,899
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/031971
; FILING DATE: 11/27/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P1014R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1994
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 121 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear

US-08-974-899-24

Alignment Scores:
Pred. No.: 111 Length: 121
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.58% Indels: 0
DB: 3 Gaps: 0

US-10-060-830-1115 (1-60) x US-08-974-899-24 (1-121)

Qy 23 GCTCAGCGTCTCGAGC 6
Db 52 AlaProAlaSerSerSer 57

RESULT 13
US-09-053-197A-23
; Sequence 23, Application US/09053197A
; Patent No. 6022952
; GENERAL INFORMATION:
; APPLICANT: Weiner, Joel H.
; APPLICANT: Turner, Raymond J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN

; TITLE OF INVENTION: SECRETION
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/053,197A
; FILING DATE: 01-APR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Macknight, Kamrin T.
; REGISTRATION NUMBER: 38,230
; REFERENCE/DOCKET NUMBER: UALB-03293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

US-09-053-197A-23

Alignment Scores:
Pred. No.: 107 Length: 158
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.58% Indels: 0
DB: 3 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-053-197A-23 (1-158)

Qy 27 CTGGGCTCCAGCGTCCTC 10
Db 98 LeuGlySerSerValLeu 103

RESULT 14
US-09-085-761A-23
; Sequence 23, Application US/09085761A
; Patent No. 6335178
; GENERAL INFORMATION:
; APPLICANT: Weiner, Joel H.
; APPLICANT: Turner, Raymond J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,761A
; FILING DATE: 28-MAY-1998

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: UALB-03356
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-085-761A-23

Alignment Scores:
Pred. No.: 107 Length: 158
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.58% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-085-761A-23 (1-158)

QY 27 CTGGGCTCCAGCGTCTC 10
|||||
Db 98 LeuGlySerValLeu 103
|

RESULT 15

US-08-606-143-45
Sequence 45, Application US/08606143
Patent No. 5836155

GENERAL INFORMATION:
APPLICANT: Li, Min
TITLE OF INVENTION: COMPOUNDS AND RELATED METHODS FOR
MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH
TITLE OF INVENTION: COMPOUNDS

NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voit & Mayer, Ltd.
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/606,143
FILING DATE: 23-FEB-1996

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kilyk Jr., John
REGISTRATION NUMBER: 30763
REFERENCE/DOCKET NUMBER: 71756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
TELEX: 25-3533

INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-606-143-45

Alignment Scores:
Pred. No.: 107 Length: 159
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.58% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-1115 (1-60) x US-08-606-143-45 (1-159)

QY 33 ACCTTGCTGGGTCCAGC 16
|||||
Db 26 ThrLeuLeuGlySerSer 31
|

Search completed: January 21, 2003, 14:53:22
Job time : 4.17323 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 21, 2003, 14:50:02 ; Search time 1.46457 Seconds
(without alignments)
1628.756 Million cell updates/sec

Title: US-10-060-830-1115
Perfect score: 20
Sequence: 1 ctgctgctcgagcagctgg.....gtggacacactgtactagcc 60

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 120991 seqs, 19878514 residues

Word size: 1

Total number of hits satisfying chosen parameters: 218080

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlh
-O=/cn2.1/USPTO.spool/US10060830/runat.16012003.092704.1550/app_query.fasta.1.1109
-DB=Published_Applications_AA -QWMT=fastan -SUFFIX=olin2p.rapp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10060830.ecgn.1.1.4 @runat.16012003.092704.1550
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:
1: /cn2.6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cn2.6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cn2.6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cn2.6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
5: /cn2.6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
6: /cn2.6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cn2.6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
8: /cn2.6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
9: /cn2.6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cn2.6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
11: /cn2.6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
12: /cn2.6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cn2.6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cn2.6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	45.0	75	10	US-09-864-761-45371
2	7	36.8	22	10	US-09-925-301-1545
3	7	36.8	347	10	US-09-098-079-14
4	7	36.8	1237	9	US-10-108-605-211

Sequence 213, App	9	US-10-108-605-213	1394	7	36.8	1394	9	US-10-108-605-213	Sequence 213, App
Sequence 41067, A	34	10	US-09-864-761-41067	6	31.6	34	10	US-09-864-761-41067	Sequence 41067, A
Sequence 41501, A	34	10	US-09-864-761-41501	6	31.6	34	10	US-09-864-761-41501	Sequence 41501, A
Sequence 331, App	8	US-08-424-5508-331	38	6	31.6	38	8	US-08-424-5508-331	Sequence 331, App
Sequence 34368, A	40	10	US-09-864-761-34368	6	31.6	40	10	US-09-864-761-34368	Sequence 34368, A
Sequence 9, Appl	56	10	US-09-945-173-9	6	31.6	56	10	US-09-945-173-9	Sequence 9, Appl
Sequence 363, App	78	9	US-09-992-598-363	6	31.6	78	9	US-09-992-598-363	Sequence 363, App
Sequence 363, App	78	9	US-09-989-293A-363	6	31.6	78	9	US-09-989-293A-363	Sequence 363, App
Sequence 66, Appl	78	9	US-10-063-547-66	6	31.6	78	9	US-10-063-547-66	Sequence 66, Appl
Sequence 15, Appl	78	9	US-09-092-296-15	6	31.6	78	9	US-09-092-296-15	Sequence 15, Appl
Sequence 363, App	78	9	US-09-989-735-363	6	31.6	78	9	US-09-989-735-363	Sequence 363, App
Sequence 363, App	78	9	US-09-990-444-363	6	31.6	78	9	US-09-990-444-363	Sequence 363, App
Sequence 363, App	78	9	US-09-989-730-363	6	31.6	78	9	US-09-989-730-363	Sequence 363, App
Sequence 363, App	78	9	US-09-991-181-363	6	31.6	78	9	US-09-991-181-363	Sequence 363, App
Sequence 363, App	78	9	US-09-993-687-363	6	31.6	78	9	US-09-993-687-363	Sequence 363, App
Sequence 363, App	78	9	US-09-989-734-363	6	31.6	78	9	US-09-989-734-363	Sequence 363, App
Sequence 363, App	78	9	US-09-997-653-363	6	31.6	78	9	US-09-997-653-363	Sequence 363, App
Sequence 242, App	78	9	US-10-174-590-242	6	31.6	78	9	US-10-174-590-242	Sequence 242, App
Sequence 242, App	78	9	US-10-176-758-242	6	31.6	78	9	US-10-176-758-242	Sequence 242, App
Sequence 363, App	78	10	US-09-989-722-363	6	31.6	78	10	US-09-989-722-363	Sequence 363, App
Sequence 363, App	78	10	US-09-989-723-363	6	31.6	78	10	US-09-989-723-363	Sequence 363, App
Sequence 363, App	78	10	US-09-989-279-363	6	31.6	78	10	US-09-989-279-363	Sequence 363, App
Sequence 363, App	78	10	US-09-989-727-363	6	31.6	78	10	US-09-989-727-363	Sequence 363, App
Sequence 363, App	78	10	US-09-989-731-363	6	31.6	78	10	US-09-989-731-363	Sequence 363, App
Sequence 363, App	78	10	US-09-991-073-363	6	31.6	78	10	US-09-991-073-363	Sequence 363, App
Sequence 363, App	78	10	US-09-990-442-363	6	31.6	78	10	US-09-990-442-363	Sequence 363, App
Sequence 363, App	78	10	US-09-991-163-363	6	31.6	78	10	US-09-991-163-363	Sequence 363, App
Sequence 363, App	78	10	US-09-993-604-363	6	31.6	78	10	US-09-993-604-363	Sequence 363, App
Sequence 363, App	78	10	US-09-990-450-363	6	31.6	78	10	US-09-990-450-363	Sequence 363, App
Sequence 363, App	78	10	US-09-989-721-363	6	31.6	78	10	US-09-989-721-363	Sequence 363, App
Sequence 242, App	78	12	US-10-006-867-66	6	31.6	78	12	US-10-006-867-66	Sequence 242, App
Sequence 46876, A	80	10	US-09-864-761-46876	6	31.6	80	10	US-09-864-761-46876	Sequence 46876, A
Sequence 30, Appl	85	10	US-09-746-491-30	6	31.6	85	10	US-09-746-491-30	Sequence 30, Appl
Sequence 116, App	88	9	US-09-774-639-116	6	31.6	88	9	US-09-774-639-116	Sequence 116, App
Sequence 33601, A	109	10	US-09-864-761-33601	6	31.6	109	10	US-09-864-761-33601	Sequence 33601, A
Sequence 47806, A	109	10	US-09-864-761-47806	6	31.6	109	10	US-09-864-761-47806	Sequence 47806, A
Sequence 120, Appl	111	10	US-09-893-737-120	6	31.6	111	10	US-09-893-737-120	Sequence 120, Appl
Sequence 14, Appl	111	10	US-09-746-491-14	6	31.6	111	10	US-09-746-491-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-09-864-761-45371 Application US/09864761
; Sequence 45371, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 45371
;; LENGTH: 75
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC013497.4
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
;; OTHER INFORMATION: EST_HUMAN HIT: A1565996.1, EVALUE 2.00e-40
;; OTHER INFORMATION: SWISSPROT HIT: P98063, EVALUE 1.00e-09
US-09-864-761-45371

Alignment Scores:
Pred. No.: 0.117 Length: 75
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 45.00% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-864-761-45371 (1-75)

Qy 34 GATGATGTGGACACTGTACTAGGC 60
Db 1 AspGlycysGlyHisThrValLeuGly 9

RESULT 2

US-09-925-301-1545
; Sequence 1545, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1545
; LENGTH: 22
; TYPE: PRT

;; ORGANISM: Homo sapiens
US-09-925-301-1545
Alignment Scores:
Pred. No.: 13.3 Length: 22
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-925-301-1545 (1-22)

Qy 45 TCCACATCCATCACCTTGCTG 25
Db 7 SerThrSerIleThrLeuLeu 13

RESULT 3

US-09-098-079-14
; Sequence 14, Application US/09098079
; Patent No. US20020064773A1
; GENERAL INFORMATION:
; APPLICANT: HerinStadt, Corrina
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Clevenger, William
; APPLICANT: Fahy, Eoin F.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON QUANTIFICATION OF
; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,079
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman Ph.D., Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 560088.416
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 347 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-098-079-14

Alignment Scores:
Pred. No.: 9.66 Length: 347
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-098-079-14 (1-347)

Qy 45 TCCACATCCATCACCTTGCTG 25
Db 1 SerThrSerIleThrLeuLeu 13

```
Db 299 SerThrsrIleThrIeuLeu 305
RESULT 4
US-10-108-605-211
; Sequence 211, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 211
; LENGTH: 1237
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-108-605-211
Alignment Scores:
Pred. No.: 8.33 Length: 1237
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 9 Gaps: 0
US-10-060-830-1115 (1-60) x US-10-108-605-211 (1-1237)
QY 23 GCTCCAGCGTCCTCGAGCAGC 3
|||||
Db 1093 AlaProAlaSerSerSer 1099
RESULT 5
US-10-108-605-213
; Sequence 213, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 213
; LENGTH: 1394
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-108-605-213
Alignment Scores:
Pred. No.: 8.21 Length: 1394
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.84% Indels: 0
DB: 9 Gaps: 0
US-10-060-830-1115 (1-60) x US-10-108-605-213 (1-1394)
QY 23 GCTCCAGCGTCCTCGAGCAGC 3
|||||
Db 1250 AlaProAlaSerSerSer 1256
RESULT 6
US-09-864-761-41067
; Sequence 41067, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 41067
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121920.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.7
```



```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-424-550B-331

Alignment Scores:
Pred. No.: 124      Length: 38
Score: 6.00      Matches: 6
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 31.58%      Indels: 0
DB: 8      Gaps: 0

US-10-060-830-1115 (1-60) x US-08-424-550B-331 (1-38)
QY 45 TCCACATCATCACCTTG 28
Db 17 SerThrSerIleThrLeu 22

RESULT 9
US-09-864-761-34368
; Sequence 34368, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica7X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34368
; LENGTH: 40

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007935.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: O88813, EVALUE 8.10e+00
; OTHER INFORMATION: EST_HUMAN HIT: AUL19344.1, EVALUE 4.00e-17
US-09-864-761-34368

Alignment Scores:
Pred. No.: 123      Length: 40
Score: 6.00      Matches: 6
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 31.58%      Indels: 0
DB: 10      Gaps: 0

US-10-060-830-1115 (1-60) x US-09-864-761-34368 (1-40)
QY 27 CTGGGCTCCAGGCTCCTC 10
Db 2 LeuGlySerValLeu 7

RESULT 10
US-09-945-173-9
; Sequence 9, Application US/09945173
; Patent No. US20020127568A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 47324, A NOVEL HUMAN G-PROTEIN AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 38155-20035.00
; CURRENT APPLICATION NUMBER: US/09/945,173
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/229,293
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid sequence
US-09-945-173-9

Alignment Scores:
Pred. No.: 119      Length: 56
Score: 6.00      Matches: 6
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 31.58%      Indels: 0
DB: 10      Gaps: 0

US-10-060-830-1115 (1-60) x US-09-945-173-9 (1-56)
QY 20 CCAGCGTCTCGAGCAGC 3
Db 48 ProAlaSerSerSerSer 53

RESULT 11
US-09-992-598-363
; Sequence 363, Application US/09992598
```

Patent No. US20020160384A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C20
CURRENT APPLICATION NUMBER: US/09/992,598
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429

; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Alignment Scores:
Pred. No.: 114
Score: 6.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 31.58%
DB: 9
Length: 78
Matches: 6
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-060-830-1115 (1-60) x US-09-992-598-363 (1-78)

QY 33 ACCTGCTGGCTCCAGC 16
|||||

Db 12 ThrLeuGlySerSer 17
|||||

RESULT 12

US-09-989-293A-363
; Sequence 363, Application US/09989293A
; Patent No. US2002017164A1
; GENERAL INFORMATION:

; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730PIC66
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04

1	PRIOR APPLICATION NUMBER: 60/088167
2	PRIOR FILING DATE: 1998-06-05
3	PRIOR APPLICATION NUMBER: 60/088202
4	PRIOR FILING DATE: 1998-06-05
5	PRIOR APPLICATION NUMBER: 60/088212
6	PRIOR FILING DATE: 1998-06-05
7	PRIOR APPLICATION NUMBER: 60/088217
8	PRIOR FILING DATE: 1998-06-05
9	PRIOR APPLICATION NUMBER: 60/088655
10	PRIOR FILING DATE: 1998-06-09
11	PRIOR APPLICATION NUMBER: 60/088734
12	PRIOR FILING DATE: 1998-06-10
13	PRIOR APPLICATION NUMBER: 60/088738
14	PRIOR FILING DATE: 1998-06-10
15	PRIOR APPLICATION NUMBER: 60/088742
16	PRIOR FILING DATE: 1998-06-10
17	PRIOR APPLICATION NUMBER: 60/088810
18	PRIOR FILING DATE: 1998-06-10
19	PRIOR APPLICATION NUMBER: 60/088824
20	PRIOR FILING DATE: 1998-06-10
21	PRIOR APPLICATION NUMBER: 60/088826
22	PRIOR FILING DATE: 1998-06-10
23	PRIOR APPLICATION NUMBER: 60/088858
24	PRIOR FILING DATE: 1998-06-11
25	PRIOR APPLICATION NUMBER: 60/088861
26	PRIOR FILING DATE: 1998-06-11
27	PRIOR APPLICATION NUMBER: 60/088876
28	PRIOR FILING DATE: 1998-06-11
29	PRIOR APPLICATION NUMBER: 60/089105
30	PRIOR FILING DATE: 1998-06-12
31	PRIOR APPLICATION NUMBER: 60/089440
32	PRIOR FILING DATE: 1998-06-16
33	PRIOR APPLICATION NUMBER: 60/089512
34	PRIOR FILING DATE: 1998-06-16
35	PRIOR APPLICATION NUMBER: 60/089514
36	PRIOR FILING DATE: 1998-06-16
37	PRIOR APPLICATION NUMBER: 60/089532
38	PRIOR FILING DATE: 1998-06-17
39	PRIOR APPLICATION NUMBER: 60/089538
40	PRIOR FILING DATE: 1998-06-17
41	PRIOR APPLICATION NUMBER: 60/089598
42	PRIOR FILING DATE: 1998-06-17
43	PRIOR APPLICATION NUMBER: 60/089599
44	PRIOR FILING DATE: 1998-06-17
45	PRIOR APPLICATION NUMBER: 60/089600
46	PRIOR FILING DATE: 1998-06-17
47	PRIOR APPLICATION NUMBER: 60/089653
48	PRIOR FILING DATE: 1998-06-17
49	PRIOR APPLICATION NUMBER: 60/089801
50	PRIOR FILING DATE: 1998-06-18
51	PRIOR APPLICATION NUMBER: 60/089907
52	PRIOR FILING DATE: 1998-06-18
53	PRIOR APPLICATION NUMBER: 60/089908
54	PRIOR FILING DATE: 1998-06-18
55	PRIOR APPLICATION NUMBER: 60/089947
56	PRIOR FILING DATE: 1998-06-19
57	PRIOR APPLICATION NUMBER: 60/089948
58	PRIOR FILING DATE: 1998-06-19
59	PRIOR APPLICATION NUMBER: 60/089952
60	PRIOR FILING DATE: 1998-06-19
61	PRIOR APPLICATION NUMBER: 60/090254
62	PRIOR FILING DATE: 1998-06-22
63	PRIOR APPLICATION NUMBER: 60/090349
64	PRIOR FILING DATE: 1998-06-23
65	PRIOR APPLICATION NUMBER: 60/090355
66	PRIOR FILING DATE: 1998-06-23
67	PRIOR APPLICATION NUMBER: 60/090431
68	PRIOR FILING DATE: 1998-06-24

```

, PRIOR FILING DATE: 1998-06-24
, PRIOR APPLICATION NUMBER: 60/090433
, PRIOR FILING DATE: 1998-06-24
, PRIOR APPLICATION NUMBER: 60/090444
, PRIOR FILING DATE: 1998-06-24
, PRIOR APPLICATION NUMBER: 60/090445
, PRIOR FILING DATE: 1998-06-24
, PRIOR APPLICATION NUMBER: 60/090472
, PRIOR FILING DATE: 1998-06-24
, PRIOR APPLICATION NUMBER: 60/090535
, PRIOR FILING DATE: 1998-06-24
, PRIOR APPLICATION NUMBER: 60/090540
, PRIOR FILING DATE: 1998-06-24
, PRIOR APPLICATION NUMBER: 60/090542
, PRIOR FILING DATE: 1998-06-24
, PRIOR APPLICATION NUMBER: 60/090557
, PRIOR FILING DATE: 1998-06-24
, PRIOR APPLICATION NUMBER: 60/090676
, PRIOR FILING DATE: 1998-06-25
, PRIOR APPLICATION NUMBER: 60/090678
, PRIOR FILING DATE: 1998-06-25
, PRIOR APPLICATION NUMBER: 60/090690
, PRIOR FILING DATE: 1998-06-25
, PRIOR APPLICATION NUMBER: 60/090694
, PRIOR FILING DATE: 1998-06-25
, PRIOR APPLICATION NUMBER: 60/090695
, PRIOR FILING DATE: 1998-06-25
, PRIOR APPLICATION NUMBER: 60/090695
, PRIOR FILING DATE: 1998-06-25
, PRIOR APPLICATION NUMBER: 60/090696
, PRIOR FILING DATE: 1998-06-25
, PRIOR APPLICATION NUMBER: 60/090862
, PRIOR FILING DATE: 1998-06-26
, PRIOR APPLICATION NUMBER: 60/090863
, PRIOR FILING DATE: 1998-06-26
, PRIOR APPLICATION NUMBER: 60/091360
, PRIOR FILING DATE: 1998-07-01
, PRIOR APPLICATION NUMBER: 60/091478
, PRIOR FILING DATE: 1998-07-02
, PRIOR APPLICATION NUMBER: 60/091544
, PRIOR FILING DATE: 1998-07-01
, PRIOR APPLICATION NUMBER: 60/091519
, PRIOR FILING DATE: 1998-07-02
, PRIOR APPLICATION NUMBER: 60/091626
, PRIOR FILING DATE: 1998-07-02
, PRIOR APPLICATION NUMBER: 60/091633
, PRIOR FILING DATE: 1998-07-02
, PRIOR APPLICATION NUMBER: 60/091978
, PRIOR FILING DATE: 1998-07-07
, PRIOR APPLICATION NUMBER: 60/091982
, PRIOR FILING DATE: 1998-07-07
, PRIOR APPLICATION NUMBER: 60/092182
, PRIOR FILING DATE: 1998-07-09

```

Alignment Scores:

Pred. No.:	114	Length:	78
Score:	6.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	31.58%	Indels:	0
DB:	9	Gaps:	0

US-10-060-830-1115 (1-60) X US-09-989-293A-363 (1-78)

Qy	33	ACCTTGCTGGGCTCCAGC	16
Db	12	ThrLeuLeuGlySerSer	17

RESULT 13

US-10-063-547-66
; Sequence 66, Application US/10063547
; Publication No. US20020182638A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen

APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3230R1C1
CURRENT APPLICATION NUMBER: US/10/063,547
CURRENT FILING DATE: 2002-05-02
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 66
LENGTH: 78
TYPE: PRT
ORGANISM: Homo Sapien
US-10-063-547-66

Alignment Scores:
Pred. No.: 114 Length: 78
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.58% Indels: 0
DB: 9 Gaps: 0

US-10-060-830-1115 (1-60) x US-10-063-547-66 (1-78)
QY 33 ACCTGCTGGCTCCAGC 16
Db 12 ThrLeuLeuGlySer 17

RESULT 14
US-09-092-296-15
Sequence 15, Application US/09092296
Publication No. US20020188114A1
GENERAL INFORMATION:
APPLICANT: BILLING-MEDEL, PATRICIA
APPLICANT: COHEN, MAURICE
APPLICANT: COLPITTS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: KLASS, MICHAEL R.
APPLICANT: RUSSELL, JOHN C.
APPLICANT: SPOUPE, STEPHEN D.
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
TITLE OF INVENTION: FOR DETECTING DISEASES OF THE LUNG
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows, Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/092,296
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/048,810
FILING DATE: 05-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6104.US.01
TELECOMMUNICATION INFORMATION:

TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 78 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. US20020188114A1e
US-09-092-296-15

Alignment Scores:
Pred. No.: 114 Length: 78
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.58% Indels: 0
DB: 9 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-092-296-15 (1-78)
QY 33 ACCTGCTGGCTCCAGC 16
Db 12 ThrLeuLeuGlySer 17

RESULT 15
US-09-989-735-363
Sequence 363, Application US/09989735
Publication No. US20020193299A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C61
CURRENT APPLICATION NUMBER: US/09/989,735
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

;
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Alignment Scores:
Pred. No.: 114 Length: 78
Score: 6.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.58% Indels: 0
DB: 9 Gaps: 0

US-10-060-830-1115 (1-60) x US-09-989-735-363 (1-78)

QY 33 ACCTTGCTGGGCTCCAGC 16
Db 12 ThrLeuGlySerSer 17
|||||

Search completed: January 21, 2003, 14:55:26
Job time : 2.46457 secs


```
QY 976 GTGGACCAATATCTATGTCTGCTGCCTACAGAAATCCCTGACAGTGTATGGGACAGAA 1035
Db 352 ThrGlnLysGlyTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1036 TGGACTGTGTACAGAGAGCTGGTGTGGAGCAAGATAGATATTTCAAGGAAACAAAGAT 1095
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysValPheGlnAlaAsnAsnAsp 389
QY 1096 TATCACAGAGTGTGGTAACTATTTTGCACCAATATTGACGTTTATTAGAGTG 1155
Db 390 AlaThrGluLeuValLeuAsnLysLeuHisThrProLeuLeuThrArgPheIleArgIle 409
QY 1156 AATCTTACCATTGGCAGCAGAAAATTCGCATGAATGGAGCTGCTCGGATGTCAGTTT 1215
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1216 ATTCTTAAGGTGCTCTCCAAACTTACTCAACCTCCACCTCCTCGGAACCAATGAC 1275
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1276 CTCAAAACACTACAGCCCTCCAAAATAGCAAGGTGCTGCCCAAAATTT----- 1329
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerThrArgGluTyrLeuTrp 460
QY 1330 -----ACGCAACCACCTACACCTCCAGTAGCAATGAAATTTCTCCACAGACAGAA 1380
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgAsnProGln 480
QY 1381 CAAACAACCTGCCAGTCTCATATCAGAAATACCTACCGTAACCTCCAAATCTAACCAAGAT 1440
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1441 GTA-----GGCTGGCTGCAGTTCTTCTCCCTGTGTG 1473
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaMetGluAlaArgAlaPhe 520
QY 1474 GTCATGGTCTCTACTCTCTATTCATATAGTGTGTGTGCTGGCAGTGG----- 1524
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1525 -----AGAAACAGAAAGAAAACCTGAAGGCACC-----TATGACTTACCTTACTGG 1572
Db 541 ProArgThrGlnGlnProLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1573 GACCGGCGAGGTTGGTGAAGAAGTAAGACAGTTTCTTCCTGCAAAAGCAGTGGACCAT 1632
Db 559 -----AspIleArgArgPhe----- 563
QY 1633 GAGGAACCCCACTTCGCTATACGACAGCGAAGTAAATCACTGAGTCCAAAGAGAGTC 1692
Db 564 GluProValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1693 ACCACAGTGTGCAGGCTGACTCTGCAGAGTAT-----GCTCAGCCACTGGTAGGAGGA 1746
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1747 ATTGTTGGTACACTTCATCAAGATCTACCTTTAAACCA-----GAAGAAGGA 1794
Db 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
QY 1795 AAGAGACGAGGCTAT-----GCAGACTAGATCCTTACAACCTCA 1833
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerGln 643
QY 1834 CCAGGGCAGGAGCTTATCATCGCTTATGCTGAACCACTCCCAATTACGGGGCCTGAGTAT 1893
Db 643 yPheAsnCysAsnPheAspPheProGluThrCysGlyTrpMetTyrAspArgAla 663
QY 1894 GCAACCCCAATCATCATGTGACATGTCCAGGCGACCCCA----- 1930
Db 663 strLeuGlnSerThrTrpIleSerSerAlaAsnProAsnAspArgThrPheProAspAs 683
```

```
QY 1931 -----CAACTTCAG-----TTGCT 1944
Db 683 pLysAsnPheLeuLysLeuGlnSerAspGlyArgGluGlyGlnPheGlyArgLeuIle 703
QY 1945 CAGCCCTCCACATCCACTTTCAAGGCTAGGGGAAACCAACCTCCCCCACTAGTGGAACT 2004
Db 703 eSerProPro-ValHisLeu-----ProArgSerProValCysMetGluP 718
QY 2005 TACAATACACTTCTCTC-----C 2022
Db 718 heGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuGlnValValArgGluAlaArg 738
QY 2023 AGGACTGACGTGCTC-----CTCAGCCAGGCCAGTATATACCCCGAAAGCTGG 2076
Db 738 InGluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlySerGluTrpLysHisGly 757
QY 2077 AAGCCAGGCTACCTGCCCCAGACGAAATTGGTGTACAGGTGCCAGACAGACACAAGAA 2136
Db 758 ArgIleIleLeuProSerTyrAsp---MetGluTyrGlnIle-----ValPheGluGly 774
QY 2137 GTATCAGGAGCAGGAAGGATGGGAATGTGATGTT 2172
Db 775 ValIleGlyLysGlyArgSerGlyGluIleSerIle 786
```

Search completed: January 21, 2003, 09:35:35
Job time : 69.9883 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 21, 2003, 09:21:35 ; Search time 37.4534 Seconds
(without alignments)
11242.484 Million cell updates/sec

Title: US-10-060-830-2
Perfect score: 3969
Sequence: 1 atgcctctgtctctctgct.....tttttaagaatactcttga 2190

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10060830/runat_16012003_091239_26859/app_query.fasta_1.6933
-DB=PIR_73 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn_1.1166 @runat_16012003_091239_26859 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	383	9.6	927	1 JQ0948	A5 antigen precurs
2	361	9.1	2211	1 KFB05	coagulation factor
3	355.5	9.0	2183	2 T42764	coagulation factor
4	351	8.8	2224	1 KFH05	coagulation factor
5	349	8.8	401	2 S65138	glycoprotein antig
6	349	8.8	427	2 S74211	PAS-6/7 protein pr
7	343	8.6	218	2 A47285	milk fat globule p
8	340	8.6	409	2 T11743	pp47 protein - pig
9	339	8.5	427	2 JC4915	ags protein precu
10	321	8.1	463	1 A36479	milk fat globule m
11	311.5	7.8	2133	2 T42763	coagulation factor
12	310	7.8	2319	2 A47004	coagulation factor
13	308.5	7.8	216	2 A44258	factor VIII-associ
14	308.5	7.8	2351	1 EZHU	coagulation factor

15	224.5	5.7	845	2 JQ5256	adipocyte transcr
16	222.5	5.6	1524	2 T30337	polyprotein - Afri
17	219.5	5.5	3623	2 T08618	intrinsic factor-B
18	215.5	5.4	3623	2 S52093	hemocytin - silkwo
19	211	5.3	3623	2 T09456	intrinsc factor-B
20	207.5	5.2	1283	2 T13799	neurexin IV - frui
21	204	5.1	719	2 S51739	transcription repr
22	200.5	5.1	855	2 S42621	hyaluronate-bindin
23	198	5.0	277	2 A41735	procollagen C-endo
24	198	5.0	707	2 JC2218	procollagen C-endo
25	195	4.9	823	1 A58788	procollagen C-endo
26	194.5	4.9	730	1 BMH01	hemin - rabbit
27	193.5	4.9	1594	2 T30549	tumor necrosis fac
28	193	4.9	275	2 JQ5506	procollagen C-endo
29	193	4.9	986	1 B58788	enteropeptidase (E
30	193	4.9	1035	1 A43090	procollagen C-endo
31	192.5	4.9	991	2 T49540	complement subcomp
32	190.5	4.8	694	2 JC6554	development protei
33	190	4.8	1464	2 S58984	complement subcomp
34	189.5	4.8	705	1 C1HURB	tolloid-BMP-1 like
35	188.5	4.7	1070	2 T31069	receptor tyrosine
36	187	4.7	913	2 A48280	TSG-6 homolog PS4
37	185.5	4.7	276	2 A47290	procollagen I C-pr
38	185.5	4.7	402	2 JH0403	complement subcomp
39	185.5	4.7	688	1 C1HUS	CRP-ductin-alpha p
40	184	4.6	2083	2 T42721	procollagen I C-pr
41	183.5	4.6	449	2 A55362	tyrosine kinase re
42	183.5	4.6	910	2 A53137	sanko - human
43	183	4.6	2403	2 A59386	dorsal-ventral pat
44	180	4.5	1057	1 A39288	oviductin (EC 3.4.
45	175	4.4	1004	2 T30338	

ALIGNMENTS

RESULT 1
JQ0948
A5 antigen precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: JH0466; JQ0948
R;Takagi, S.; Hirata, T.; Agata, K.; Mochii, M.; Eguchi, G.; Fujisawa, H.
Neuron 7, 295-307, 1991
A;Title: The A5 antigen, a candidate for the neuronal recognition molecule, has homol
A;Reference number: JH0466; MUID:91337458; PMID:1908252
A;Accession: JH0466
A;Molecule type: mRNA
A;Residues: 1-927 <TAK>
A;Cross-references: GB:D10467; GB:D01077; NID:G222962; PIDN:BAA01260.1; PID:G222963
A;Experimental source: tadpole, brain
A;Note: This protein has motifs homologous to complement components C1r and C1s and t
C;Comment: This protein is a neuronal cell surface molecule involved in the neuronal
C;Superfamily: Xenopus A5 antigen; C1r/C1s repeat homology; discoidin I amino-termina
C;Keywords: duplication; glycoprotein; transmembrane protein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-927/Product: A5 antigen #status predicted <A5A>
F;27-138/Domain: C1r/C1s repeat homology <C1r1>
F;147-262/Domain: C1r/C1s repeat homology <C1r2>
F;274-424/Domain: discoidin I amino-terminal homology <DNI>
F;430-584/Domain: discoidin I amino-terminal homology <DN2>
F;646-812/Domain: MAM homology <MAM>
F;861-883/Domain: transmembrane #status predicted <TM>
F;150,261,300,523,844/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 1.18e-20
Score: 383.00
Length: 927
Matches: 138
Conservative: 70
Percent Similarity: 38.88%
Best Local Similarity: 25.79%
Query Match: 9.63%
Mismatches: 197
Indels: 130
Gaps: 22
DB:

US-10-060-830-2 (1-2190) x JQ0948 (1-927)

QY 13 CTCCTCTCTACTTGTCTC-----CTGCTCTCTCTCTCTCTGCTGAGGACGCTGGA 57
 Db 1 MetLeuLeuArgLeuLeuSerCysCysTyrProSerGlnArgCysSerLeuArgSerSerTyr 20
 QY 58 GCCACAGAGGTGATGGATGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTTACA 117
 Db 21 AlaSerArgAsnAspLysCysGlyAspThrIleLysIleThrSerProSerTyrLeuThr 40
 QY 118 TCATTAACATCCACAGACCTATCCCAACAGCTGTTGTGTGAATGGAGAGTCCGTGA 177
 Db 41 SerAlaGlyTyrProHisSerTyrProSerGlnArgCysGlyTyrLeuIleGlnAla 60
 QY 178 ---AGATGGGAGAGAGATTCGCAATTTGTG---GACTTTCACATGAAGATTCT 231
 Db 61 ProGluHisTyrGlnArgIleMetIleAsnPheAsnProHisPheAspLeuGluAspArg 80
 QY 232 GATTCCTGTCACTTTAAATTTACTTGAATTTATATGAATTTGAGTGGAGTCAAGTCAA 291
 Db 81 Glu---CysLysTyrAspTyrValGluValIleAspGlyAspAsnAlaAsnGlyGlnLeu 99
 QY 292 ATAGCAAAATCTGTGGTCTGGGTGGCAATGAACCAATTCA---ATTGAATCAAAAGGC 348
 Db 100 LeuGlyLysTyrCysGly-----LysIleAlaProSerProLeuValSerThrGly 116
 QY 349 AATGAATCACAATCTCTGATGAGTGGAAATCCATGTTCTGACGCGGATTTTGGCC 408
 Db 117 ProSerIlePheIleArgPheValSerAspTyrGluThrProGlyAlaGlyPheSerIle 136
 QY 409 TCATACTCTGTTATAGATAAACAAGATCTAATTACTTGTGTTGGACACTGCATCCAATTT 468
 Db 137 ArgTyrGluValPheLys----- 142
 QY 469 TTGGAACCTGAGTTCACTAGTACTGCCAGCTGGTTGTCTCTCTCTCTCTGAGATA 528
 Db 143 ThrGlyProGlu-----CysSerArgAsnPheThrSerSer 154
 QY 529 TCTGGAACAATT-----CCTCATGGATATAGATATCC-----TCG 564
 Db 155 AsnGlyValIleLysSerProLysTyrProGluLysTyrProAsnAlaLeuGluCysThr 174
 QY 565 CCATTGTGCATGCTGTGTCATGCAGGAGTAGTG----- 600
 Db 175 TyrlleIlePheAlaProLysMetGlnGluIleValLeuGluPheGluSerPheGluLeu 194
 QY 601 -----TCAACAGCTTGGGCGGCCAAATCAGT-----GTTGTAATTAGT 639
 Db 195 GluAlaAspSerAsnAlaProGlyGlyGlnThrCysArgTyrAspTyrPueGlyIleTyr 214
 QY 640 AAAGGTATTCCC-----TATTATGAAAGTTCTTTGGCTAAC 675
 Db 215 AspGlyPheProGlyValGlyProHisIleGlyArgTyrCysGlyGlnAsnThrProGly 734
 QY 676 AAGCTCATCTCTGGTGGGACACTTATCTACACAGTCTTTTACATTTAAGACAAGTGA 735
 Db 235 ArgValArgSerPheThrGlyIleLeuSerMetIlePheHisThrAspSerAlaIleAla 735
 QY 735 ----- 735
 Db 255 LysGluGlyPhePheAlaAsnPheSerValValGlnSerAsnThrAspGluAspPheGln 274
 QY 736 TGTTATGGAACACTGGGATGGAGTCTGTGTGATCGCGGATCCTCAATAACAGCATCA 795
 Db 275 CysLysGluAlaLeuGlyMetGluSerGlyGlyIleHisPheAspGlnIleSerValSer 294
 QY 796 TCT-----GTGCTGGAGTGGAGTACCACACAGGGCAA-----CAGACAGT 837
 Db 295 SerGlnTyrSerMetAsnTyrSerAlaGluArgSerArgLeuAsnTyrValGluAsnGly 314
 QY 838 TGAACACCAAAAAAGCCAGCTGAAACAACTGGACCGCTTGGCTGGCTTTTCCACT 897
 Db 315 TrpThrProGlyGluAspThrVallys----- 323

QY 898 GATGAATACCACTGCTTACAAATAGATTTGAATAAGAAAAGAAATAACAGGCATTATA 957
 Db 324 -----GluTrpIleGlnValAspLeuGluAsnLeuArgPheValSerGlyIleGly 340
 QY 958 ACCACTGGA-----TCCACCATGGTGGAGCACAAATTACTATGTCTGCTCCCTACACAAATC 1011
 Db 341 ThrGlnGlyAlaIleSerLysGluThrLysLysLysTyrPheValLysSerTyrLysVal 360
 QY 1012 CTGTACACTGATGATGGCGAGAAATGGACTGTGTACAGAGAGCCGTGTGGAGCAAGAT 1071
 Db 361 AspIleSerSerAsnGlyGluAspTyrIleThrLeuLysAspGly-----AsnLysHis 378
 QY 1072 AAGATATTTCAGGAACAAAGATTATCACCAGGATGTGCGTAATAACTTTTGGCCACA 1131
 Db 379 LeuValPheThrGlyAsnThrAspAlaThrAspValValTyrArgProPheSerLysPro 398
 QY 1132 ATTATTGACACTTTTATTAGAGTGAATCCTACCCAATGGCAGCAGAAATTCCTCATGAA 1191
 Db 399 ValIleThrArgPheValArgLeuArgProValThrTrpGluAsnGlyIleSerLeuArg 418
 QY 1192 ATGAGCTGCTCGGACGATGTCAGTTTATTCTCTAAAGTCTGCTCCCAAACTTACTCAACCT 1251
 Db 419 PheGluLeuTyrGlyCys-----LysIleThrAspTyr 429
 QY 1252 CCACCTCTCTCGGAACAGCAATGACCTC-----AAAACACTACAGCC 1293
 Db 430 ProCysSerArgMetLeuGlyMetValSerGlyLeuIleSerAspSerGlnIleThrAla 449
 QY 1294 CCTCCAAAATAGCAAGGTCGTGCCCAAAATTTACGCAACCACTACACCTCGCAGT 1353
 Db 450 SerSerGlnValAspArgAsnTrpValProGluLeuAlaArgLeuValThrSerArgSer 469
 QY 1354 -----AGCAATGAATTTCTCTGCACAGACAGAA 1380
 Db 470 GlyTrpAlaLeuProProSerAsnThrHisProTyrThrLysGlu 484
 RESULT 2
 KE905
 coagulation factor V precursor - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
 C:Accession: A42580; A36497
 R:Quinto, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.
 J. Biol. Chem. 267, 2971-2978, 1992
 A:Title: The complete cDNA sequence of bovine coagulation factor V.
 A:Reference number: A42580; MUID:92147638; PMID:1737753
 A:Accession: A42580
 A:Molecule type: mRNA
 A:Residues: 1-2211 <GUI>
 A:Cross-references: GB:M81440; NID:G163037; PIDN:AAA30512.1; PID:G163038
 A:Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBI:80776)
 R:Kalafatis, M.; Jenny, R.J.; Mann, K.G.
 J. Biol. Chem. 265, 21580-21589, 1990
 A:Title: Identification and characterization of a phospholipid-binding site of bovine
 A:Reference number: A36497; MUID:91072354; PMID:2254316
 A:Accession: A36497
 A:Molecule type: protein
 A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584; 1673-1676, 'X', 1678-1679, 'X', 1681, 'X'
 R:Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.
 Biochemistry 33, 13109-13116, 1994
 A:Title: Determination of the disulfide bridges in factor Va heavy chain.
 A:Reference number: A55979; MUID:95034740; PMID:7947716
 A:Contents: annotation
 A:Note: 566-Cys and 617-Cys were shown to have free sulphydryls
 C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
 C:Function:
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa p
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
 C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasm
 F:1-28/Domains: signal sequence #status predicted <SIG>
 F:29-2211/Product: coagulation factor V #status predicted <MAT>
 F:29-741/Product: coagulation factor Va heavy chain #status predicted <VAH>

Qy	691	GTGGGACACTTACTACAGTCTTTTACATTATAGACAAAGTGGATGCTTATGGACACTG	750
Db	2013	-----LeuArgLeuGluLeuGlnGlyCysGluValAsnGlyCysSerThrProLeu	2029
Qy	751	GGGATGGAGTCTGGTGTGATCGGGATCTCAATAAACAGCATCATCTGTGCTGGAG--	807
Db	2030	GlyLeuGluAspGlyArgIleGlnAspGlyGlnIleThrAlaSerSerPheLysLysSer	2049
Qy	808	---TGGACTGACCACAGGCGGCAAGAGACAGATTGGAAACCCCAAAAGCCAGCGCTGAAA	864
Db	2050	TriPrpGlyAspTyr-----TrrGluProSerLeuAlaArgLeuAsn	2063
Qy	865	AAACCTGGA-----CGGCTTGGGCTGCTTTGGCCACTGATGAATACCAGTGGTTACAA	918
Db	2064	AlaGlnGlyArgValAsnAlaTrpGlnAlaLysAlaAsnAsnAsnLysGlnTrpLeuGln	2083
Qy	919	ATAGATTGTAATAGGNAAGAAATAACAGCATTATAACCACTGGATCCACCATGGTG	978
Db	2084	ValAspLeuLeuLysIleLysLysValThrAlaIleValThrGlnGlyCysLysSerLeu	2103
Qy	979	GAGCACAAATTACTATGCTGTGCTGCTACAGAACTCTGTACAGTGTATGGCAGAAATGG	1038
Db	2104	SerSerGluMetTyrValLysSerTyrSerIleGlnTyrSerAspGlnGlyValAlaTrp	2123
Qy	1039	ACTGTGTACAGAGAGCGCTGTGTGGAGCAAGATAGATATTTCAAGGNAACAAGATTAT	1098
Db	2124	LysProTyrArgGlnLysSerSerMetValAspLysIlePheGluGlyAsnSerAsnThr	2143
Qy	1099	CACAGGATGCGGTATAACTTTTGGCCACCAATATATGCAGCTTTATTAGTAGTGAAT	1158
Db	2144	LysGlyHisMetLysAsnPhePheAsnProIleIleSerArgPheIleArgIleLe	2163
Qy	1159	CCTACCCAATGGCAGCAGAAAAATTGCCATGAATAAGGAGCTGCTCGGATGT	1209
Db	2164	ProLysThrTrpAsnGlnSerIleAlaLeuArgLeuGluLeuPheGlyCys	2180
RESULT 4			
KFH05			
coagulation factor V precursor [validated] - human			

KFHU5
coagulation factor V precursor [validated] - human

F:1906-2061/Domain: discoidin I amino-terminal homology <DN1>
F:2065-2224/Domain: C2 <DC2>
F:2065-2221/Domain: discoidin I amino-terminal homology <DN2>
F:51_55_239_297_460_468_554_741_752_760_776_782_821_938_977_1074_1083_1103_1106_1479_1573-1574/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experimental
F:167-193_248-329_500-526_603-684_1725-1751_1907-2061_2066-2221/Disulfide bonds: #sta
F:334-335/cleavage site: Arg-Asn (protein C) #status predicted
F:363_693_1346/Binding site: sulfate (tyr) (covalent) #status predicted
F:376-377/cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:382_1338/Binding site: carbonylate (Asn) (covalent) #status absent
F:534-535/cleavage site: Arg-Gly (protein C) #status predicted
F:737-738/cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experimental
F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Alignment Scores:
Pred. No.: 3 42e-18 Length: 2224
Score: 35.00 Matches: 77
Percent Similarity: 59.04% Conservative: 21
Best Local Similarity: 46.39% Mismatches: 58
Query Match: 8.84% Indels: 10
DB: 1 Gaps: 3

US-10-060-830-2 (1-2190) x KFHU5 (1-2224)

QY 724 AAGCAAGTGGATGTTATGAACACTGGGGATGGAGTCTGGTGATCGCGGATCCCTCAA 783
:: :::::::::::::::::::: ||| |||
Db 2062 GluValAsnGlyCysSerThrProLeuGlyMetGluasnglyLysIleGluasnlsgln 2081
:: :::::::::::::::::::: ||| |||
QY 784 ATAACACATCATCTGTCTGGAG-----TGGACTGACCACACAGCGGCAAGAACAGT 837
||||||| ||| :: ::
Db 2082 IleThrAlaSerSerPheLysLysSerTrpTrpGlyAspTyr----- 2095
:: :::::::::::::::::::: ||| |||
QY 838 TGGAAACCCAAAAGCCAGCGCTGAAAAACCTGGA-----CGGCCTTGGGCTGCTTTT 891
||||| :::::::::::::::::::: ||| |||
Db 2096 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2115
:: :::::::::::::::::::: ||| |||
QY 892 GCCACTGATGAATACCACTGGTGTACAAATAGATTGAATAAGAAAAGAAAATAACAGGC 951
||| :::::::::::::::::::: ||| |||
Db 2116 AlaAsnAsnLysGlnTrpLeuGluIleAspleuLysIleLysIleThrAla 2135
||| :::::::::::::::::::: ||| |||
QY 952 ATTATACCACTGGATCCCACTGGTGGGACACAATTACTATGTCTGCCTCAGAATC 1011
||||||| ||| :: ::
Db 2136 IleIleThrGlnGlyCysLysSerLeuSerSerGluMetTyrValLysSeryThrIle 2155
||||||| ||| :: ::
QY 1012 CTGTACAGTCATGATGGCCAGAAATGGACTGTGTACAGAGAGCGCTGGTGGAGCAAGAT 1071
||||||| :::::::::::::::::::: ||| |||
Db 2156 HisTyrSerGluGlnGlyValGluTrpLysProTyrArgLeuLysSerSerMetValasp 2175
||| :::::::::::::::::::: ||| |||
QY 1072 AGATATTTCAGNAACAAGATTATCACAGAGATGTGCGTAATAACTTTTTGCCACCA 1131
||| :::::::::::::::::::: ||| |||
Db 2176 LysIlePheGluGlyAsnThrAsnThrLysGlyHisValAsnLysPheAsnProPro 2195
||| :::::::::::::::::::: ||| |||
QY 1132 ATATTGACGTTTTATTAGATGATCTACCCATGGCAGCAAAAATTGCCATGAA 1191
||||||| :::::::::::::::::::: ||| |||
Db 2196 IleIleSerArgPheIleargValIleProLysThrTrpAsnGlnSerIleAlaLeuAraG 2215
||| :::::::::::::::::::: ||| |||
QY 1192 ATGGAGTCCTGGGATGT 1209
||||||| :::::::::::::::::::: ||| |||
Db 2216 LeuGluLeuPheGlyCys 2221
||| :::::::::::::::::::: ||| |||

RESULT 5
S5138
glycoprotein antigen MGP57/53, mammary gland - bovine (fragment)
N:Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E
C:Species: Bos primigenius taurus (cattle)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
R:Accession: S5138; G48394
A:AOkl, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A>Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal
A:Reference number: S5138; MUID:96125736; PMID:9541316
A:Accession: S5138
A>Status: preliminary; not compared with conceptual translation

A; Molecule type: mRNA

A; Residues: 1-401 <AOK>

R; Author: I.H.; Banghart, L.R.; Lane, W.S.

Biochem. Mol. Biol. Int. 29, 545-554, 1993

A; Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig II-like sequences.

A; Reference number: A48394; MUID:93250576; PMID:8485470

A; Accession: G48394

A; Status: preliminary

A; Molecule type: protein

A; Residues: 207-220 <MAT>

A; Experimental source: milk

A; Note: sequence extracted from NCBI backbone (NCBIP:131457)

C; Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homology

C; Keywords: glycoprotein

F; 1-32/Domain: EGF homology (fragment) <EG1>

F; 40-79/Domain: EGF homology <EG2>

F; 82-239/Domain: discoidin I amino-terminal homology <DN1>

F; 243-401/Domain: discoidin I amino-terminal homology <DN2>

Alignment Scores:

Pred. No.:	3,7e-18	Length:	401
Score:	349.00	Matches:	77
Percent Similarity:	56.59%	Conservative:	26
Best Local Similarity:	42.31%	Mismatches:	68
Query Match:	8.79%	Indels:	11
DB:	2	Gaps:	5

US-10-060-830-2 (1-2190) x S65138 (1-401)

Qy 684 ATCTGT-----GGTGGGACACTTATCTACAGTCTTTTACATTAAAGCAAGTGATG 737

Db 225 ILeCysHisArgGlyCysThrLeuArgPheGlu-LeuLeuGlyCysGluLeuAsnGlyCy 244

Qy 738 TTATGGACACTGGGATGGAGTCTGCTGTCGCGATCCCTCAATAACAGCATCATC 797

Db 244 sThrGluProLeuGlyLeuLysAspAsnThrIleProAsnLysGlnIleThrAlaSerSe 264

Qy 798 TGTGCTGGAG---TGGACTGACCCACACAGGG---CAAGAGAACAGATTGGAAACCAAAA 851

Db 264 rTyTyTyThrTrp-----GlyLeuSerAlaPheSerTrpPheProTyTy 280

Qy 852 AGCCAGGCTGAAAAACCTGGACCG-----CCTTGGCTGCTTTTGGCCACTGATGAATA 905

Db 280 rAlaArgLeuAspAsnGlnGlyLysPheAsnAlaTrpThrAlaGlnThrAsnSerAlaSe 300

Qy 906 CCAGTGTTCACAAATAGATTGAATAAGAAAGAAATAACAGGCATTATAACCACTGG 965

Db 300 rGluTrpLeuGlnIleAspLeuGlySerGlnLysArgValThrGlyIleIleThrGlnG 320

Qy 966 ATCCACCATGTTGGAGCAACAATTACTATGTCTGCTGCCCTACAGATPCCTGTACAGTGATGA 1025

Db 320 yAlaArgAspPheGlyHisIleGlnTyValAlaAlaTyArgValAlaTyGlyAspAs 340

Qy 1026 TGGGAGAAATGGACTGTGTACAGAGACCTGGTGGAGCAAGATAGATATTCACAGG 1085

Db 340 pGlyValThrTrpThrGluTyLysAspProGlyAlaSerGluSerLysIlePheProGl 360

Qy 1086 AAACAAACATATCACAGGATGTCGCTTAATAACTTTTGGCCACCAATTATTGACAGTTT 1145

Db 360 yAsnMetAspAsnAsnSerHisLysLysAsnIlePheGluThrProPheGlnAlaArg 380

Qy 1146 TATTAGTGAATCTTACCAATGCGACAGAAATATGCGATGAAATGGAGCTGCTCGG 1205

Db 380 eValArgIleGlnProValAlaTrpHisAsnArgIleThrLeuArgValGluLeuGln 400

Qy 1206 ATGT 1209

Db 400 yCys 401

RESULT 6

S74211

PAS-6/7 protein precursor - bovine

C; Species: Bos primigenius taurus (cattle)

C; Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000

C; Accession: S74211; S78114; S24181

R; Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.

Eur. J. Biochem. 240, 628-636, 1996

A; Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat g

A; Reference number: S74211; MUID:97008954; PMID:8856064

A; Accession: S74211

A; Molecule type: mRNA

A; Residues: 1-427 <HVA>

A; Cross-references: EMBL:X91895; NID:gl632778; PIDN:CAA62997.1; PID:gl632779

A; Accession: S78114

A; Molecule type: protein

A; Residues: 19-85; 96-110; 140-165; 174-216; 221-232; 248-277; 285-293; 309-337; 339-420; 425-

R; Kim, D.H.; Kanno, C.; Mizokami, Y.

Biochim. Biophys. Acta 1122, 203-211, 1992

A; Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, f

A; Reference number: S23926; MUID:92353107; PMID:1643094

A; Accession: S24181

A; Molecule type: protein

A; Residues: 383-394 <KIM>

C; Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom

C; Keywords: blocked amino end; disulfide bond; glycoprotein; milk

F; 1-18/Domain: signal sequence #status predicted <SIG>

F; 19-427/Product: PAS-6/7 protein #status experimental <MAT>

F; 24-58/Domain: EGF homology <EG1>

F; 66-105/Domain: EGF homology <EG2>

F; 108-265/Domain: discoidin I amino-terminal homology <DN1>

F; 269-427/Domain: discoidin I amino-terminal homology <DN2>

F; 24-35, 29-47, 49-58, 66-77, 71-94, 96-105/Disulfide bonds: #status predicted

F; 27/Binding site: carbohydrate (Ser) (covalent) #status experimental

F; 34/Binding site: carbohydrate (Thr) (covalent) #status experimental

F; 59, 227/Binding site: carbohydrate (Asn) (covalent) #status experimental

F; 109-265, 252-256, 270-427/Disulfide bonds: #status experimental

Alignment Scores:

Pred. No.:	3,73e-18	Length:	427
Score:	349.00	Matches:	77
Percent Similarity:	56.59%	Conservative:	26
Best Local Similarity:	42.31%	Mismatches:	68
Query Match:	8.79%	Indels:	11
DB:	2	Gaps:	5

US-10-060-830-2 (1-2190) x S74211 (1-427)

Qy 684 ATCTGT-----GGTGGGACACTTATCTACAGTCTTTTACATTAAAGCAAGTGATG 737

Db 251 ILeCysHisArgGlyCysThrLeuArgPheGlu-LeuLeuGlyCysGluLeuAsnGlyCy 270

Qy 738 TTATGGACACTGGGATGGAGTCTGCTGTCGCGATCCCTCAATAACAGCATCATC 797

Db 270 sThrGluProLeuGlyLeuLysAspAsnThrIleProAsnLysGlnIleThrAlaSerSe 290

Qy 798 TGTGCTGGAG---TGGACTGACCCACACAGGG---CAAGAGAACAGATTGGAAACCAAAA 851

Db 290 rTyTyTyThrTrp-----GlyLeuSerAlaPheSerTrpPheProTyTy 306

Qy 852 AGCCAGGCTGAAAAACCTGGACCG-----CCTTGGCTGCTTTTGGCCACTGATGAATA 905

Db 306 rAlaArgLeuAspAsnGlnGlyLysPheAsnAlaTrpThrAlaGlnThrAsnSerAlaSe 326

Qy 906 CCAGTGTTCACAAATAGATTGAATAAGAAAGAAATAACAGGCATTATAACCACTGG 965

Db 326 rGluTrpLeuGlnIleAspLeuGlySerGlnLysArgValThrGlyIleIleThrGlnG 346

Qy 966 ATCCACCATGTTGGAGCAACAATTACTATGTCTGCTGCCCTACAGATPCCTGTACAGTGATGA 1025

Db 346 yAlaArgAspPheGlyHisIleGlnTyValAlaAlaTyArgValAlaTyGlyAspAs 366

Qy 1026 TGGGAGAAATGGACTGTGTACAGAGACCTGGTGGAGCAAGATAGATATTCACAGG 1085

Db 366 pGlyValThrTrpThrGluTyLysAspProGlyAlaSerGluSerLysIlePheProGl 386

Qy	1086	AAACAAGATTATCACCGAGTGTGGTAATAACTTTTGGCCACCAATTAATGCACTTT	1143
Db	386	yAsnMetAspAsnSerHisLysLysAsnIlePheGluThrProPheGlnAlaArgPh	406
Qy	1146	TATTAGAGTGAACTACCACCATGGCAGACAAAATGCCATGAATGGAGCTGCTCGG	1205
Db	406	eValArgIleGlnProValalaTrpHisAsnArgileThrLeuargValGluLeuLeuGl	426
Qy	1206	ATGT 1209	
Db	426	yCys 427	
RESULT 7			
A47285			
milk fat globule protein - human (fragment)			
C:Species: Homo sapiens (man)			
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999			
C:Accession: A47285			
R:Larocca, D.; Peterson, J.A.; Urrea, R.; Kuniyoshi, J.; Bistrain, A.M.; C			
Cancer Res 51, 4994-4998, 1991			
A>Title: A Mr 46,000 human milk fat globule protein that is highly expressed			
A:Reference number: A47285; MOID:91371351; PMID:1909932			
A:Accession: A47285			
A>Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 1-218 <LAR>			
A:Cross-references: GB:S56151; NID:9235396; PIDN:AAB19771.1; PID:g235397			
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology			
F:1-56/Domain: discoidin I amino-terminal homology (fragment) <DN1>			
F:60-218/Domain: discoidin I amino-terminal homology <DN2>			
Alignment Scores:			
Pred. No.:	9,49e-18	Length:	218
Score:	343.00	Matches:	71
Percent Similarity:	57.23%	Conservative:	28
Best Local Similarity:	41.04%	Mismatches:	68
Query Match:	8.64%	Indels:	6
DB:	2	Gaps:	3
US-10-060-830-2 (1-2190) x A47285 (1-218)			
Qy	700	TTATCTCAAGTCTTTTTACATTTAAGACAAGTGGATGTTATGAACACTGGGATGGAG	759
Db	49	LeuArgPheGluLeuLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys	68
Qy	760	TCGTGTGTATCCGGATCCTCAAATAACAGCATCATCTGTCTGGAG---TGCACTGAC	816
Db	69	AsnAsnSerIleProAspLysGlnIleThrAlaSerSerSeryrTyLysThrrTrpGlyLeu	88
Qy	817	CACACAGGCGACAGAACAGTGTGGAAACCCAAAAAGCCAGCGCTGAAAAAACCTGGGA---	873
Db	89	HisLeu-----PheSerTrpAsnProSerTyrrAlaArgLeuAspLysGlnGlyAsn	105
Qy	874	--CCGCCTTGGCTGCTTTTGGCACTGATGAATACCACTGGTTACAAATAGATTGGAAT	930
Db	106	PheAsnAlaTrpValAlaGlySeryrGlyAsnAspGlnTrpLeuGlnValAspLeuGly	125
Qy	931	AAGCAAGAAGAAATACAGGCATTATACCACTGGATCCACCATGGTGGGACCACAATAC	990
Db	126	SerSerLysGluValThrGlyIleIleThrGlnGlyAlaArgAsnPheGlySerValGln	145
Qy	991	TATGTGTCGCCACAGAATCCCTGACAGTAGATGGCGAGAATGGACTGTGTACAGA	1050
Db	146	PheValAlaSerTyrrLysValalaTyrrSerAsnAspSerAlaAsnTrpThrGluTrpGln	165
Qy	1051	GAGCTCGTGTGGACAGATAGATATTTCGAAGGAAACAAGATTATFCACCGAGTGTG	1110
Db	166	AspProArghThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisSerHisLys	185
Qy	1111	CGTAATAACTTTTGGCCACCATTAATGCACTTTTATAGATGAATCCTACCAATGG	1170
Db	186	LysAsnLeuPheGluThrProfileuAlaArgTyrrValrgIleLeuProValAlaTrp	205

[illegible]

408 vCvs 409

RESULT 9

JC4915
 ags protein precursor - rat
 N;Alternate names: O-acetyl-Gd3 ganglioside
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
 C;Accession: JC4915
 R;Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
 Biochem. Biophys. Res. Commun. 225, 932-938, 1996
 A;Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
 A;Reference number: JC4915; MUID:96374422; PMID:8780713
 A;Accession: JC4915
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-427 <QUG>
 A;Cross-references: DDBJ:DR84068; NID:g1620006; PIDN:BAA12210.1; PID:g1620007
 A;Experimental source: CST cell
 C;Comment: This protein is required for the O-acetylation of disialoganglioside sialic acid.
 C;Genetics:

A:Gene: ags
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homology
F:1-21/Domain: signal sequence #status predicted <SIG>
F:28-60/Domain: EGF homology <EG1>
F:68-107/Domain: EGF homology <EG2>
F:110-267/Domain: discoidin I amino-terminal homology <DN1>
F:271-427/Domain: discoidin I amino-terminal homology <DN2>

Alignment Scores:		
Pred. No.:	2,41e-17	Length:
Score:	339.00	Matches:
Percent Similarity:	38.66%	Conservative:
Best Local Similarity:	26.89%	Mismatches:
Query Match:	8.54%	Indels:
DB:	2	Gaps:
		19
		427

US-10-060-830-2 (1-2190) x JC4915 (1-427)

QY	13	CTCTGCTCTTACTTGTCCCTGCTCTGCTGCTGAGGACCTGGAGCCACGACAGGTGAT	72
Db	7	LeuAlaIaLeucySGlyValLeuLeucySAlaSerGlyLeuPheAlaIaSerGlyAsp	26
QY	73	GGATGCT	96
Db	27	PheCysAspSerSerLeuCysLeuAsnGlyGlyThrCysLeuMetGlyGlnAspAsnAsp	46
QY	97	-----CCTGAGATGGAAACCTTTACATCCATAAATACCACCCAGACCTTAT	141
Db	47	IleItyrCysLeucyProGluGlyPheThrGlyLeuValCysAsnGluThrGluIbysGly	66
QY	142	CCCAACAGC---ACTGTTTGTGAATGGGAGATCCGTAAAGATGGGACGAGAGATTGGC	198
Db	67	ProCysSerProAsnProCysPheHisAspAlaIySLeuValThrGluAspThr---	85
QY	199	ATCAAAATTTGGTACTTGTGACATTGAAGATCTCGATTCTGTTCATCTTAATTA	258
Db	86	---GlnArgGlyAspIlePheThrGluItyrIleCysGlnCysProValGlyTyrSerGly	104
QY	259	ATTTTATATGAATTTGGAGTCACGACAGACACTGAATAGGC	297
Db	105	IleHisCysGluLeuGlyCysSer---ThrIysLeuGlyLeuGluGlyAlaIleAla	123
QY	298	-----AAATACTGTGGT---CTGGGGTTGCAATGAACACCAT	330
Db	124	AspSerGlnIleSerAlaSerSerValTyrMetGlyPheMetGlyLeuGln	140
QY	331	TCAAATTGAATCAAAAGGCAATGAATACATTCGCTGTTTCATGAGTGGATTC	381
Db	141	-----ArgTrpGlyProGluLeuAlaArgLeuItyrArgThrGlyIleValAsnAla	157
QY	381	-----	381

RESULTS 10

RESULT 10
 A36479
 milk fat globule membrane protein - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C;Accession: A36479
 R;Stubbs, J.D.; Lekutis, C.; Singer, K.L.; Bui, A.; Yuzukui, D.; Srinivasan, Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
 A;Title: cDNA cloning of a mouse mammary epithelial cell surface protein re
 A;Reference number: A36479; MUID:91046008; PMID:2122462
 A;Accession: A36479

A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-463 <STU>
 A:Cross-references: GB:M30337; NID:gl919142; PIDN:AA39534.1; PID:gl919143
 C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homology
 C:Keywords: membrane protein
 F:28-60/Domain: EGF homology <EG2>
 F:68-107/Domain: EGF homology
 F:147-303/Domain: discoidin I amino-terminal homology <DN1>
 F:307-463/Domain: discoidin I amino-terminal homology <DN2>

Alignment Scores:
 Pred. No.: 4, 8e-16 Length: 463
 Score: 321.00 Matches: 63
 Percent Similarity: 57.76% Conservative: 30
 Best Local Similarity: 39.13% Mismatches: 62
 Query Match: 8.09% Indels: 6
 DB: 1 Gaps: 3

US-10-060-830-2 (1-2190) x A36479 (1-463)

QY 733 GGATGTTATGAACACTGGGATGAGTCTGTGTGATGCGGATCCCTCAATAACAGCA 792
 DB 307 GlyCysLeuGluProLeuGlyLeuLysAsnAsnThrIleProAspSerGlnMetSerAla 326
 QY 793 TCATCTGTCTGGAGTGGACTGACCACACAGGCGCAAGAGACAGTGGAAACCCAAAAA 852
 DB 327 SerSer-----SerTyrIlyThrTrpAsnLeuArgAlaPheGlyTyrTrpProHisLeu 344
 QY 853 GCCAGCTGAAACAACTGGACCG-----CCTGGGTGCTTTTGGCCACTGATGAATAC 906
 DB 345 GlyArgLeuAspAsnGlnGlyLysIleAsnAlaTrpThrAlaGlnSerAsnSerAlaLys 364
 QY 907 CAGTGGTTACAAATAGATTGAATGAAGAAAGAAATAACAGGCAATATAACCACTGA 966
 DB 365 GluTrpLeuGlnValAspLeuGlyThrGlnArgGlnValThrGlyIleIleThrGlnGly 384
 QY 967 TCCACCATGGTGAGCAACAAATACTATGTCTGCTCAGCAATCCTGTACAGTGATGAT 1026
 DB 385 AlaArgAspPheGlyHisIleGlnTyrValGluSerTyrLysValAlaHisSerAsp 404
 QY 1027 GGGCAGAAATGACGTGTACAGAGAGCGTGTGTGGAGCAAGATAAGATATTCACGA 1086
 DB 405 GlyValGlnTrpThrValTyrGluGlnGly-----SerSerLysValPheGlnGly 422
 QY 1087 ACAAAAGATTATCACAGATGCGTAACTACTTTTCCACCAATATTGCCAGCTTTT 1146
 DB 423 AsnLeuAspAsnAsnSerHisLysLysAsnIlePheGluLysProPheMetAlaArgTyr 442
 QY 1147 ATTAGATGAATCCTACCAATGGCAGCAGAAATTTGCCATGGAATGGAGTGTGTCGA 1206
 DB 443 ValArgValLeuProValSerTrpHisAsnArgIleThrLeuArgLeuGluLeuLeuGly 462
 QY 1207 TGT 1209
 DB 463 Cys 463

RESULT 11
 T42763
 coagulation factor VIII precursor - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
 C:Accession: T42763
 R:Lollar, P.
 submitted to the EMBL Data Library, August 1996
 A:Reference number: Z22269
 A:Accession: T42763
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2133 <LOL>
 A:Cross-references: EMBL:U49517; NID:gl511633; PID:gl511634; PIDN:AA06705.1
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
 C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla

F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-2133/Product: coagulation factor VIII #status predicted <MAT>
 F:23-349/Domain: ferroxidase repeat homology <FOX1>
 F:402-730/Domain: ferroxidase repeat homology <FOX2>
 F:1498-1820/Domain: ferroxidase repeat homology <FOX>

Alignment Scores:
 Pred. No.: 3, 15e-15 Length: 2133
 Score: 311.50 Matches: 126
 Percent Similarity: 38.04% Conservative: 68
 Best Local Similarity: 24.71% Mismatches: 174
 Query Match: 7.85% Indels: 142
 DB: 2 Gaps: 19

US-10-060-830-2 (1-2190) x T42763 (1-2133)

QY 9 GTTCTCTGCTTACTTGTCTGCT-----CCTGCTGCTCGAGACGC----- 53
 DB 1653 IleGlyProLeuLeuIleCysArgAlaAsnThrLeuAsnAlaAlaHisGlyArgGlnVal 1672
 QY 54 -----TGGAGCCCAAGAGTGTGATGTGG 80
 DB 1673 ThrValGlnGluPheAlaLeuPheThrIlePheAspGluThrLysSerTrpTyrPhe 1692
 QY 81 ACA-----CACTGTACTAGCCCT-----GAGAGTGAAC 110
 DB 1693 ThrGluAsnValGluArgAsnCysArg-AlaProCysHisLeuGlnMetGluAspProth 1712
 QY 111 CTTTACATCCATAACTACCCACAGACACCTATCCCAACAGCAGCTTTGTGTAATGGGAGAT 170
 DB 1712 rLeuLysGlu---AsnTyrArgPheHisAlaIleAsnGlyTyrValMetAspThrLeuPr 1731
 QY 171 CCGTAAAGATGGGAGAGAGAGTTCGCATCAAAATTTGGTGACTTTGACATT----- 222
 DB 1731 oGlyLeuValMetAlaGlnAsnGlnArgIleArgTrpTyrLeuLeuSerMetGlySerAs 1751
 QY 223 -GAATCTCTGATCTTGTCTCACTTTAAT----- 249
 DB 1751 nGluAsnIleHisSerIleHisPheSerGlyHisValPheSerValArgLysGluGlu 1771
 QY 250 ----TACTTGAGAAATTTAAT----- 267
 DB 1771 uTyrLysMetAlaValTyrAsnLeuTyrProGlyValPheGluThrValGluMetLeuPr 1791
 QY 268 ----GGAATTTGGATGACAGCAAGTAA-----ATAGCAAAATCTGTGTCTGGGTT 317
 DB 1791 oSerLysValGlyIleTrpArgIleGluCysLeuIleGlyGluHisLeuGlnAlaGlyMe 1811
 QY 318 GCAATGAACCATTCATTTGAATCAAAAGCAATGAAATCACATTCCTCTTCATGAGTGG 377
 DB 1811 tSerThrThrPheLeuValTyrSerLysGluCysGlnAlaProLeuGlyMetAlaSerGlu 1831
 QY 378 AATCCATCTTTCTGGACGCGGATTTTGGCCCTCATCTCTGTTATAGATAAACAAGATCT 437
 DB 1831 y-----ArgIleArgAspPheGln----- 1837
 QY 438 AATTACTTTGTTGGACACTGCATCCAAATTTTGGAACTGAGTTCAGTAAGTACTGCC 497
 DB 1838 -----IleThrAlaSerGly-----GlnTyrGlyGlnTrpAlaPr 1849
 QY 498 AGCTGGTTGTCTGCTTCCTTTTGGCTGAGATATCTGNAACAAT----- 540
 DB 1849 oLysLeuAlaArgLeuHisTyr-----SerGlySerIleAsnAlaTrpSerThrLy 1866
 QY 541 ----CCTCATGGATATAGAGATTCCTCGCCATTTGTCATGGCTGGTGTGTGTCGAGGAGT 596
 DB 1866 sAspProHisSerTrpIleLysValAspLeuLeuAlaProMetIleIleHisGlyIleMe 1886
 QY 597 AGTGTCAACACGTTGGCGGCCAAATCAGTGTGTGTAATTAGTAAGATGATTCCTATTA 656
 DB 1886 tThrGlnGlyAlaArgGlnLysPheSerSerLeuTyrIleSerGlnPheIleIleMetTy 1906
 QY 657 TGA----- 660

Db 1906 rSerLeuaspClyArgasnTrpGlnSerTyrArgGlyAsnSerThrGlyThrLeuMetVa 1926
 Qy 661 -----AGTTCTTTGGCTTAACAACGTC----- 681
 Db 1926 lPhePheGlyAsnValaspAlaSerGlyIleLysHisAsnIlePheAsnProIleVa 1946
 Qy 682 -----ACATCTGTGGTGGACACTTATCTACAA 710
 Db 1946 lAlaArgTyrIleArgLeuHisProThrHisTyrSerIleArgSerThrLeuArgMetGl 1966
 Qy 711 TCTTTTACATTAAGCAACAGGTATGATGAACACATGGGAGTGGTGTGTGAT 770
 Db 1966 uLeuMetClyCysaspLeuAsnSerCysSerMetProLeuGlyMetGlnAsnLysAlaIl 1986
 Qy 771 CGCGATCTCAATAACAGCATCATCTGTCTGGAGTGGACTGACACACAGGCGCAAGA 830
 Db 1986 eSerAspSerGlnIleThrAlaSerSerHisLeuSerAsnIlePheAlaThr----- 2003
 Qy 831 GAACAGTTGGAACCCAAAGCCAGGCTGAAGAAACCTGGA-----CCGCCCTTGGGC 884
 Db 2004 -----TrpSerProSerGlnAlaArgLeuHisLeuGlnGlyArgThrAsnAlaTrpAr 2021
 Qy 885 TGCTTTTCCCATGATGAATACCATGGTGTCAAAATAGATTGTAAGGAAAGAAAT 944
 Db 2021 gProArgValSerSerAlaGluTrpLeuGlnValaspLeuGlnLysThrValLysVa 2041
 Qy 945 AACAGGCATTATACCATGGATCCACCATGGTGGAGCACAAATTAATGTCTGCGCTA 1004
 Db 2041 lThrGlyIleThrGlnGlyValLysSerLeuSerSerMetTyrValLysGluPh 2061
 Qy 1005 CAGATCTGTACATGATGATGGCAGCAATGGACTGTGTACAGAGACCCGTGGTGA 1064
 Db 2061 eLeuValSerSerSerGlnAspGlyArgArgTrpThrLeuPheGlnAspGly----- 2079
 Qy 1065 GCAAGATAGATATTAAAGAAACAAAGATTATCACCAGGATGTGCGCTAAATAACTTTT 1124
 Db 2080 -HisThrLysValPheGlnGlyAsnGlnAspSerSerThrProValValAsnAlaLeuAs 2099
 Qy 1125 GCCACCAATTATGACGTTTATAGAGTGAATCCTACCCAAATGGCAGCAGCAAAATGCG 1184
 Db 2099 pProLeuPheThrArgTyrLeuArgIleHisProThrSerTrpAlaGlnHisIleAl 2119
 Qy 1185 CATGAAATGGAGTCGCGGATGTCAG 1212
 Db 2119 aleuArgLeuGluValLeuGlyCysGlu 2128
 RESULT 12
 A47004
 C:Species: Mus musculus (house mouse)
 C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 18-Jun-1999
 C:Accession: A47004
 R:Elder, B.; Lakich, D.; Gitschler, J.
 Genomics 16, 374-379, 1993
 A:Title: Sequence of the murine factor VIII CDNA.
 A:Reference number: A47004; MUID:93300511; PMID:8314577
 A:Accession: A47004
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2319 <ELD>
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:23-349/Domain: ferroxidase repeat homology <FO>
 F:402-730/Domain: ferroxidase repeat homology <FO2>
 F:1686-2006/Domain: ferroxidase repeat homology <FO3>
 F:2007-2156/Domain: discoidin I amino-terminal homology <DNI>
 F:2160-2313/Domain: discoidin I amino-terminal homology <DN2>
 Alignment Scores:
 Pred. No.: 4,14e-15 Length: 2319
 Score: 310.00 Matches: 96

Percent Similarity: 44.76% Conservative: 45
 Best Local Similarity: 30.48% Mismatches: 113
 Query Match: 7.81% Indels: 61
 DB: 2 Gaps: 11
 US-10-060-830-2 (1-2190) x A47004 (1-2319)
 Qy 306 TGGTCTGGGTTGCAAAATGAACCAATTCAATTGAATCAAAAGCAATGAATCAACATTGCT 365
 Db 2049 TrpSer-----ThrLysGluProPheSerTrpIle-LysValaspLeuAlaProMe 2066
 Qy 366 GTTCATGAGTGAATCCCATCTTTCTGGA-----CG 395
 Db 2066 tIleValHisGlyIleLysThrGlnGlyAlaArgGlnLysPheSerSerLeuTyrIleSe 2086
 Qy 396 CGGATTTTGGCCCTCATCTCTGTATATAGATAAACAACATTAATTACTTGTGTCACAC 455
 Db 2086 rGlnPheIleIleMetTyrSerLeuaspGlyLysLys-----TrpLeuSerTy 2102
 Qy 456 TGCATCCCAATTTTGTGGAACCTGAGTTCAGTAAGTACTGCCAGCTGCTGTCTGCTTCC 515
 Db 2102 rGlnGlyAsn-----SerThrGlyThrLeuMetVa 2112
 Qy 516 TTTTGTCTGAGATATCTGGAACAATTCCTCATGGATATAGAGATTCTCCGCCATTGTGAT 575
 Db 2112 lPhe-----PheGlyAsnValaspSerSer----- 2120
 Qy 576 GGCTGGTGTGCATGAGGAGTAGTCTCAACACAGTTGGCGGCCCAATCAGTGTGTGAAT 635
 Db 2121 -----GlyIleLysHisAsnSerPheAsnProIle-----Il 2132
 Qy 636 TAGTAAGGTATTCCTTATTGAAGTTCCTTGGCTAACACGTCACATCTGTGTGGG 695
 Db 2132 eAlaArgTyrIleArgLeuHisProThrHis-----SerSerIleArgSe 2147
 Qy 696 ACACATATCTACACGCTTTTACATTTAAGACAAAGTGTATGGAACACTGGGAT 755
 Db 2147 rThrLeuArgMetGluLeuMetGlyCysaspLeuAsnSerCysSerIleProLeuGlyMe 2167
 Qy 756 GGAGTCTGGTGTGTCGGGATCCTCAATAACACATCATCTGTGTGGAGTGGACGAC 815
 Db 2167 tGluSerLysValIleSerAspThrGlnIleThrAlaSerSerTyrPhe----- 2183
 Qy 816 CCACACAGGGCAAGAGAACAGATTGGAAACCCAAAGCCAGGCTGAAGAAACCTGGA-- 873
 Db 2184 -----ThrAsnMetPheAlaThrTrpSerProSerGlnAlaArgLeuHisLeuGlnGlyAr 2202
 Qy 874 ----CCGCCCTTGGGCTGCTTTTGGCCACTGATGAATACCACTGGTGTACAAATAGATTGAA 929
 Db 2202 gThrAsnAlaTrpArgProGlnValAsnAspProLysGlnTrpLeuGlnValaspLeuGl 2222
 Qy 930 TAAGGAAAGAAATAACAGGATATATACCACTGATCCACCATGGTGGAGCACAATTA 989
 Db 2222 nLysThrMetLysValThrGlyIleIleThrGlnGlyValLysSerLeuPheThrSerMe 2242
 Qy 990 CTATGTCTGCTACAGAAATCCTGTACAGATGATGGCAGAAATGACGTGTGTACAG 1049
 Db 2242 tPheValLysGluPheLeuLysSerSerGlnaspGlyHisHisTrpThrGlnIleLe 2262
 Qy 1050 AGAGCCTGGTGGAGCAAGATAAGATATTTCAAGAAACAAGATATTCACCAAGATGT 1109
 Db 2262 uTyrAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSerSerThrProMe 2280
 Qy 1110 GGTATTAATCTTTTCCCACTTATTCACAGTTTATAGATGTAATCCCTACCAATG 1169
 Db 2280 tMetAsnSerLeuaspProLeuLeuThrArgTyrLeuArgIleHisProGlnIleTr 2300
 Qy 1170 GCAGCAGAAATTTGCCATGAAATGAGTGTGCTCGGATGTCAG 1212
 Db 2300 pGluHisGlnIleAlaLeuArgLeuGluIleLeuGlyCysGlu 2314
 RESULT 13
 A44258

factor VIII-associated gene B hypothetical protein - human

C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C:Accession: A44258

R:Levinson, B.; Kenwick, S.; Gamel, P.; Fisher, K.; Gitschler, J.

Genomics 14, 585-589, 1992

A:Title: Evidence for a third transcript from the human factor VIII gene.

A:Reference number: A44258; MUID:93052386; PMID:1427887

A:Accession: A44258

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-216 <LEV>

A:Cross-references: GB:M90707; NID:g182316; PIDN:AAA58466.1; PID:g182317

C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase

F:1-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>

F:57-210/Domain: discoidin I amino-terminal homology <DN2>

Alignment Scores:

Pred. No.:	3,71e-15	Length:	216
Score:	308.50	Matches:	66
Percent Similarity:	54.49%	Conservative:	31
Best Local Similarity:	37.08%	Mismatches:	72
Query Match:	7.77%	Indels:	9
DB:	2	Gaps:	3

US-10-060-830-2 (1-2190) x A44258 (1-216)

QY	685	TCTGTGGGACACTTATCTACAGTCTTTTACATTTAAGACAAGTGGTGTATGGA	744
Db	41	: LeuArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet	60
QY	745	ACACTGGGGTGGAGTGTGTGTGTCGCGGATCCTCAATAACAGCATCATCTGTGCTG	804
Db	61	: LeuMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyrPhe	80
QY	805	GAGTGGACTACACACAGGGCAAGACAGAGTGGGAAACCCAAAGCCAGCGCTGAAA	864
Db	81	: ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis	95
QY	865	AAACCTGGA-----CGCCCTGGGCTGCTTTGCCACTGATGAATACCGTGGTTACAA	918
Db	96	LeuGlnGlyArgSerAsnAlaThrPargProGlnValAsnAsnProLysGluTrpLeuGln	115
QY	919	ATAGATTGTAATAGGAAAGAAATAACAGGCATTATAACCATCGTGCATCCATGGTG	978
Db	116	ValAspPheGlnLysThrMetLysValThrGlyValThrGlnGlyValLysSerLeu	135
QY	979	GAGCACAAATTAATGTGTCTGCTACAGAACTCTGACAGTGTATGGGCGAGAAATGG	1038
Db	136	LeuThrSerMetTyrValLysGluPheLeuIleSerSerGlnAspGlyHisGlnTrp	155
QY	1039	ACTGTGTACAGAGCCCTGGTGTGGACAGATAGATATTTCAAGGAACAAGATATAT	1098
Db	156	ThrLeuPhePheGlnAsnGly-----LysValLysValPheGlnGlnGlnAspSer	173
QY	1099	CACCAGGATGTGCGTATAAATATTTTCCACCAATTTATGCGCTTTTATAGAGTGAAT	1158
Db	174	PheThrProValAsnSerLeuAspProLeuLeuThrArgTyrLeuArgIleHis	193
QY	1159	CCTACCCAAATCGCAGAGAAATTTGCCATGAAATGAGGTGCTCGGATGTCAG	1212
Db	194	ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu	211
RESULT	14		
E2HU			
		coagulation factor VIII precursor [validated] - human	
		N:Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant compo	
		C:Species: Homo sapiens (man)	
		C>Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 08-Dec-2000	
		C:Accession: 154318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445; B42	
		R:Gitschler, J.; Wood, W.I.	
		Hum. Mol. Genet. 1, 199-200, 1992	
		A:Title: Sequence of the exon-containing regions of the human factor VIII gene.	

A:Reference number: I54318; MUID:93265012; PMID:1303178

A:Accession: I54318

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-1921, 'S', 1923-2351 <RES>

A:Cross-references: GB:M88648; NID:g182381; PIDN:AAA52420.1; PID:g182383

R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschler, J.; Keyt, B.; See

Nature 312, 330-337, 1984

A:Title: Expression of active human factor VIII from recombinant DNA clones.

A:Reference number: A00525; MUID:85061548; PMID:6438526

A:Accession: A00525

A:Molecule type: mRNA

A:Residues: 1-2351 <WOOS>

A:Cross-references: EMBL:X01166; EMBL:X01179

R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.

S, D.N.; Hewick, R.M.

Nature 312, 342-347, 1984

A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.

A:Reference number: I58059; MUID:85061550; PMID:6438528

A:Accession: I58059

A>Status: nucleic acid sequence not shown; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-74, 'V', 76-1259, 'E', 1261-2351 <RE2>

A:Cross-references: GB:K01740; NID:g182802; PIDN:AAA52484.1; PID:g182803

R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.;

B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favaloro, J.; Hansen, J.

DNA 4, 333-349, 1985

A:Title: Characterization of the polypeptide composition of human factor VIII:C and

A:Reference number: A23584; MUID:86081164; PMID:3935400

A:Accession: A23584

A:Molecule type: mRNA

A:Residues: 1-2351 <TRU>

A:Cross-references: GB:M4113; NID:g182817; PIDN:AAA52485.1; PID:g182818

R:Eaton, D.; Rodriguez, H.; Vehar, G.A.

Biochemistry 25, 505-512, 1986

A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleavag

ity.

A:Reference number: A26174; MUID:86159740; PMID:3082357

A:Accession: A26174

A:Molecule type: protein

A:Residues: 20-36;332-399, 'X', 401-402;1668-1678;1709-1722, 'D', 1723-1725;1741-1755 <E

R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.

Biochemistry 31, 3315-3325, 1992

A:Title: Identification and functional importance of tyrosine sulfate residues withi

A:Reference number: A42348; MUID:92207952; PMID:1554716

A:Accession: A42348

A:Molecule type: protein

A:Residues: 20-36;336-371;392-408;582-594;1668-1669, 'X', 1671;1672-1692;1693-1708;170

A:Experimental source: recombinant material from Chinese hamster ovary cells

A:Note: sequence extracted from NCBI backbone and corrected to correspond with the p

R:Pay, P.J.; Smudzin, T.M.

J. Biol. Chem. 264, 14005-14010, 1989

A:Title: Intersubunit fluorescence energy transfer in human factor VIII.

A:Reference number: A43986; MUID:89340500; PMID:2503509

A:Accession: A43986

A:Molecule type: protein

A:Residues: 'X', 517-523;1853-1860, 'X', 1862-1864, 'X', 1866 <FAY>

R:Leyte, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens,

J. Biol. Chem. 266, 740-746, 1991

A:Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential

A:Reference number: A56109; MUID:91093266; PMID:1898735

A:Contents: annotation; sulfation

R:Gitschler, J.; Wood, W.I.; Goralaka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Veh

Nature 312, 326-330, 1984

A:Title: Characterization of the human factor VIII gene.

A:Reference number: A56196; MUID:85061547; PMID:6438525

A:Contents: annotation; introns

R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.

Protein Sci. 4, 740-746, 1995

A:Title: Locations of disulfide bonds and free cysteines in the heavy and light chai

A:Reference number: A56216; MUID:95338127; PMID:7613471

A:Contents: annotation; disulfide bonds

A:Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls

R;Kjalke, M.; Heding, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
 Eur. J. Biochem 234, 773-779, 1995
 A:Title: Amino acid residues 721-729 are required for full factor VIII activity.
 A:Reference number: S63527; MUID:96163459; PMID:8575434
 A:Accession: S63527
 A:Molecule type: protein
 A:Residues: 733-752;753-759 <KJA>
 R;Lind, P.; Larsson, K.; Spira, J.; Sydow-Baeckman, M.; Almstedt, A.; Gray, E.; Sandberg
 Eur. J. Biochem 232, 19-27, 1995
 A:Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
 A:Reference number: S66445; MUID:96048024; PMID:7556150
 A:Accession: S66445
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1668-1685 <LIN>
 C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure pr
 C:Genetics:
 A:Gene: GDB:F8C
 A:Cross-references: GDB:119124; OMIM:306700
 A:Map position: Xq28-Xq28
 A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
 C:Function:
 A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor VIII; discoidein I amino-terminal homology; ferroxidase
 C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-2351/Product: coagulation factor VIII #status experimental <MAT>
 F:20-740/Product: coagulation factor VIIIA heavy chain #status experimental <ACH>
 F:20-356/Domain: A1 <DA1>
 F:23-348/Domain: ferroxidase repeat homology <F01>
 F:392-759/Domain: A2 <DA2>
 F:402-730/Domain: ferroxidase repeat homology <F02>
 F:760-1667/Domain: B <DB0>
 F:1668-2351/Product: coagulation factor VIIIA light chain #status experimental <ACL>
 F:1709-2038/Domain: A3 <DA3>
 F:1716-2038/Domain: ferroxidase repeat homology <F03>
 F:2039-2191/Domain: C1 <DC1>
 F:2039-2188/Domain: discoidein I amino-terminal homology <DN1>
 F:2192-2351/Domain: C2 <DC2>
 F:2192-2351/Domain: discoidein I amino-terminal homology <DN2>
 F:160-258,601,776,803,847,919,962,1020,1074,1085,1204,1274,1278,1301,1319,1403,1
 F:172-198,267-348,547-730,1851-1877,1918-1922,2040-2188/Disulfide bonds: #status
 F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted
 F:365,737,738,742,1683,1699/Binding site: sulfate (Tyr) (covalent) #status experimental
 F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:414,426/Binding site: sulfate (Tyr) (covalent) #status predicted
 F:759-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental
 F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental
 F:2193-2345/Disulfide bonds: #status predicted

Alignment Scores:
 Pred. No.: 5,38e-15 Length: 2351
 Score: 308.50 Matches: 66
 Percent Similarity: 54.49% Conservative: 31
 Best Local Similarity: 37.08% Mismatches: 72
 Query Match: 7.77% Indels: 9
 DB: 1 Gaps: 3

US-10-060-830-2 (1-2190) x EZHU (1-2351)

QY 685 TCTGTGGGGACACTATCTACAGCTCTTTTACATTTTAAGCAAGCTGGATGTATGGA 744
 |||||
 Db 2176 SerIleArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet 2195
 QY 745 ACATCGGGATGGAGTGGTGTGATCGCGATCCTCAATACACAGCATCATCTGTGCTG 804
 |||||
 Db 2196 ProLeuGlyMetGluSerIleAlaIleSerAspAlaGlnIleThrAlaSerSerTyrPhe 2215
 QY 805 GAGTGGACTGACACACAGGGCAAGAGACAGTTGGAAACCCCAAAAGCCAGCGTGAA 864
 |||||

Db 2216 -----ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis 2230
 QY 865 AAACCTGGA-----CGCCCTGTGGCTGCTTTTCCCACTGATGAATACCACTGGTTACAA 918
 |||||
 Db 2231 LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGln 2250
 QY 919 ATGATTGTAATAGGAAAGAAATAACAGGATATTAACCACTGCATCCACCATGGTG 978
 |||||
 Db 2251 ValAspPheGlnLysThrMetLysValThrGlyValThrThrGlnGlyValLysSerLeu 2270
 QY 979 GAGCACAAATACATGCTGCTGCTCAGACATCTGTACAGTGTATGATGGCAGAAATGG 1038
 |||||
 Db 2271 LeuThrSerMetTyrValLysGluPheLeuIleSerSerGlnAspGlyHisGlnTrp 2290
 QY 1039 ACTGTGTACAGAGAGCCGTGTGGACAGATAAGATATTTCAAGAAACAAGATTAT 1098
 |||||
 Db 2291 ThrLeuPhePheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer 2308
 QY 1099 CACCAGGATGTGCTTAATACCTTTTCCCACTATTTTGCACCAATATTATAGAGTGAAT 1158
 |||||
 Db 2309 PheThrProValValAsnSerLeuAspProLeuLeuThrArgTyrLeuArgIleHis 2328
 QY 1159 CCTACCAATGGCAGCAGAAATATGCTGATGAAATGCGATGCTGCGGATGTCAG 1212
 |||||
 Db 2329 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 2346
 RESULT 15
 JC5256
 A:adipocyte transcription factor, AEBP1 - human
 C:Species: Homo sapiens (man)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: JC5256
 R;Ohno, I.; Hashimoto, J.; Shimizu, K.; Takaoka, K.; Ochi, T.; Matsubara, K.; Okubo,
 Biochem. Biophys. Res. Commun. 228, 411-414, 1996
 A:Title: A cDNA cloning of human AEBP1 from primary cultured osteoblasts and its expr
 A:Reference number: JC5256; MUID:97075196; PMID:8920928
 A:Accession: JC5256
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-845 <OHN>
 A:Cross-references: DDBJ:D86479; NID:g1468942; PIDN:BAAL3094.1; PID:g1468943

Alignment Scores:
 Pred. No.: 9.38e-09 Length: 845
 Score: 224.50 Matches: 52
 Percent Similarity: 47.90% Conservative: 28
 Best Local Similarity: 31.14% Mismatches: 68
 Query Match: 5.66% Indels: 19
 DB: 2 Gaps: 4

US-10-060-830-2 (1-2190) x JC5256 (1-845)

QY 748 CTGGGATGGAGTCTGGTGTGATCGGGATCCTCAATAACAGCATCATCTGTGCTGGAG 807
 |||||
 Db 75 IleGlyMetGluSerHisArgIleGluAspAsnGlnIleArgAlaSerSerMetLeu--- 93
 QY 808 TGGACTGACCACACAGGCGAAGACAGTTCGAACCCCAAAAGCCAGGCTGAAAAA 867
 |||||
 Db 94 -----ArgHisGlyLeuGlyAlaGlnArgGlyArgLeuAsnMet 106
 QY 868 CCT-----GGACCCCTTGGGCTGCTTTTCCCACTGAT 900
 |||||
 Db 107 GlnThrGlyAlaThrGluAspAspTyrTyrAspGlyAlaTrpCysAlaGluAspAla 126
 QY 901 GAATACCACTGGTTACAAATAGATTGTAAGGAAAGAAATAACAGCATTAATATAACC 960
 |||||
 Db 127 ArgThrGlnTrpIleGluValAspThrArgArgThrThrArgPheThrGlyValIleThr 146
 QY 961 ACTGGATCCCACTGGTGGAGGACCAATTAATCTGTCTGCCTACAGATTCCTGTACAGT 1020
 |||||
 Db 147 GlnGlyArgAspSerSerIleHisAspAspPheValThrThrPhePheValGlyPheSer 166
 QY 1021 GATGATGGCAGAAATGGACTGTGTACAGAGACCTGGTGTGGAGACCTGTGTGATAGATATTT 1080

```
Db 167 AsnAspSerGlnThrTrpValMetThrAsn---GlyTyrGlu---GluMetThrPhe 184
Qy 1081 CAAGGAAACAAAGATTATCACCAGGATGCGGTAACTTTTGGCCACCAATTATTGCA 1140
Db 185 HisGlyAsnValAspLysAspThrProValLeuSerGluLeuProGluProValValAla 204
Qy 1141 CGTTTATTAGATGAATCTTACCCTTACCCCAATGGCAGCAGAAATTTGCCATCAAAATGGAGCTG 1200
Db 205 ArgPheIleArgIleTyrProLeuThrTrpAsnGlySerLeuCysMetArgLeuGluVal 224
Qy 1201 CTCGGATGTCAGTTTATTCCT 1221
Db 225 LeuGlyCysSerValAlaPro 231
```

Search completed: January 21, 2003, 09:41:36
Job time : 55.4534 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 07:26:10 ; Search time 17.3647 Seconds
(without alignments)
10461.794 Million cell updates/sec

Title: US-10-060-830-2
Perfect score: 3969
Sequence: 1 atgcctctgttcctctgct.....ttttaagaatacctttga 2190

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-Q/cn2_1/USPTO.spool/US10060830/runat_16012003_091238_26841/app_query.fasta_1.6933
-DB=SwissProt_40 -Qfmt=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830_6CGN_1_1_71@runat_16012003_091238_26841 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	421.5	10.6	925	1 NRP2_RAT	Q35276 rattus norv
2	412.5	10.4	931	1 NRP2_MOUSE	Q35375 mus musculu
3	400.5	10.1	923	1 NRPL_HUMAN	O14786 homo sapien
4	398.5	10.0	931	1 NRP2_MOUSE	O60462 homo sapien
5	391.5	9.9	923	1 NRPL_MOUSE	P97333 mus musculu
6	387.5	9.8	922	1 NRPL_RAT	Q9qwt9 rattus norv
7	384.5	9.7	914	1 NRPL_CHICK	P79795 gallus gall
8	383	9.6	928	1 NRPL_XENLA	P28824 xenopus lae
9	362	9.1	2258	1 FA5_PIG	Q9gip1 sus scrofa
10	361	9.1	2211	1 FA5_BOVIN	Q28107 bos taurus
11	349	8.8	427	1 MFGM_BOVIN	Q95114 bos taurus
12	347	8.7	2224	1 FA5_HUMAN	P12259 homo sapien
13	343	8.6	387	1 MFGM_HUMAN	Q08431 homo sapien
14	340	8.6	409	1 MFGM_PIG	P79385 sus scrofa
15	339	8.5	427	1 MFGM_RAT	P70490 rattus norv
16	321	8.1	463	1 MFGM_MOUSE	P21956 mus musculu
17	311.5	7.8	2133	1 FA8_PIG	P12263 sus scrofa
18	310	7.8	2319	1 FA8_MOUSE	Q06194 mus musculu

19	308.5	7.8	2351	1	FA8_HUMAN	P00451 homo sapien
20	269.5	6.8	224	1	XLRI_MOUSE	Q92114 mus musculu
21	261.5	6.6	224	1	XLRI_HUMAN	O15537 homo sapien
22	256.5	6.5	280	1	XLRI_FUGRU	Q946r5 fugu rubrip
23	217.5	5.5	1331	1	CTA2_HUMAN	Q9uhc6 homo sapien
24	215.5	5.4	3133	1	HMCT_BOMMO	P98092 bombyx mori
25	208	5.2	1310	1	CTA4_MOUSE	Q99p47 mus musculu
26	207.5	5.2	1284	1	NRX4_DROME	Q94887 drosophila
27	206	5.2	1308	1	CTA4_HUMAN	Q9c0a0 homo sapien
28	202	5.1	1022	1	TLD_BRARE	O57460 brachydanio
29	200.5	5.1	855	1	DDR2_HUMAN	Q16832 homo sapien
30	198	5.0	277	1	TSG6_HUMAN	P98066 homo sapien
31	198	5.0	707	1	BMP1_XENLA	P98070 xenopus lae
32	193	4.9	275	1	TSG6_MOUSE	O08859 mus musculu
33	193	4.9	986	1	BMP1_HUMAN	P13497 homo sapien
34	193	4.9	1035	1	ENTK_BOVIN	P98072 bos taurus
35	192.5	4.9	991	1	BMP1_MOUSE	P98063 mus musculu
36	189.5	4.8	705	1	CIR_HUMAN	P00736 homo sapien
37	188.5	4.7	911	1	DDR1_MOUSE	Q03146 mus musculu
38	186	4.7	468	1	PCOL1_RAT	O08628 rattus norv
39	186	4.7	913	1	DDR1_HUMAN	Q08345 h epithelia
40	185.5	4.7	276	1	TSG6_RABIT	P98065 oryctolagus
41	185.5	4.7	688	1	C1S_HUMAN	P09871 homo sapien
42	184.5	4.6	468	1	PCOL1_MOUSE	Q61398 mus musculu
43	184.5	4.6	639	1	BMPH_STRPU	P98069 strongyloce
44	184	4.6	854	1	DDR2_MOUSE	Q62371 mus musculu
45	183.5	4.6	910	1	DDR1_RAT	Q63474 rattus norv

ALIGNMENTS

RESULT 1

ID	NRP2_RAT	STANDARD;	PRT;	925 AA.
AC	Q35276;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DE	Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).			
GN	NRP2.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RA	Kolodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J.,			
RA	Ginty D.D.;			
RT	Neuropilin is a semaphorin III receptor."			
RL	Cell 90:753-762(1997).			
CC	-!- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165 AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- TISSUE SPECIFICITY: FOUND IN CERTAIN NEURONAL POPULATIONS OF THE CNS AND IN OTHER NONNEURONAL TISSUES INCLUDING MESENCHYMAL TISSUE LINING IN THE RIBS.			
CC	-!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.			
CC	-!- SIMILARITY: CONTAINS 2 CUB DOMAINS.			
CC	-!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.			
CC	-!- SIMILARITY: CONTAINS 1 MAM DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF016297; AAC53338.1; -			

DR HSP: P12259; 1CZT.
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000421; FA58_C.
 DR InterPro: IPR000998; MAM_domain.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00629; MAM; 1.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00231; FA58C; 2.
 DR SMART: SM00137; MAM; 1.
 DR PROSITE: PS01180; CUB; 2.
 DR PROSITE: PS01285; FA58C_1; 2.
 DR PROSITE: PS01286; FA58C_2; 2.
 DR PROSITE: PS50060; MAM_2; 1.
 KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor.
 FT SIGNAL 1 22
 FT CHAIN 23 925
 FT DOMAIN 23 858
 FT TRANSMEM 859 883
 FT DOMAIN 884 925
 FT DOMAIN 28 142
 FT DOMAIN 149 267
 FT DOMAIN 277 427
 FT DOMAIN 434 592
 FT DOMAIN 642 802
 FT DISULFID 28 55
 FT DISULFID 83 105
 FT DISULFID 149 175
 FT DISULFID 208 230
 FT DISULFID 277 427
 FT DISULFID 434 592
 FT CARBOHYD 152 152
 FT CARBOHYD 157 157
 FT CARBOHYD 629 629
 FT CARBOHYD 833 833
 FT CARBOHYD 834 834
 SQ SEQUENCE 925 AA; 103896 MW; 3BF62903P644851C CRC64;

Alignment Scores:
 Pred. No.: 6.12e-23 Length: 925
 Score: 421.50 Matches: 201
 Percent Similarity: 36.42% Conservative: 110
 Best Local Similarity: 23.54% Mismatches: 347
 Query Match: 10.62% Indels: 196
 DB: 1 Gaps: 34

US-10-060-830-2 (1-2190) x NRP2_RAT (1-925)

QY 1 ATGCTCTGTTCCTCCCTCTTACTGTCTGTCTGTCTGTCTGCTC-----GAGGAGGCT 54
 DB 1 MetaspMetPheProLeuThrTrpIlePheLeuAlaLeuTyrPheSerGlyHisLysVal 20
 QY 55 GGAGCCGAGCAAGGTGATGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTT 114
 DB 21 ArgSerGlnGlnAspProProCysGlyGlyArgLeuAsnSerLysAspAlaGlyTyrIle 40
 QY 115 ACATCCATAACTACCCACAGACCTATCCCAACAGCACTCTTGTGAATGGAGATCCGT 174
 DB 41 ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluTrpValTyr 60
 QY 175 GTA---AAGATGGGAGAGAGTCCGATCAAAATTTGGT---GACITTTGACATGAGAT 228
 DB 61 AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGluLys 80
 QY 229 TCTGATCTTCTCACTTAATTAATCTGAGAAATTAATGGAATTTGGAGTCAGCAGAACT 288
 DB 81 Hisasp---CysLysTyrAspPheIleGluIleArgAspGlyAspSerGluSerAlaAsp 99
 QY 289 GAAATAGCAAAATCTGTGTGGGTTCGCAATGAAC-----CATTCAAATGAA 339
 DB 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProThrIleIle 114
 QY 340 TCAAAGGCAATGAATCAATTCATTCCTGCTTCATGATGGGAATTCATGTTCTTGTGACCGGGA 399

DB 115 SerSerGlySerValLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly 134
 QY 400 TTTTGGCCTCATCTCTGTATAGATAAACAAGATTAATTAATCTTCTTTGGACACTGCA 459
 DB 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
 QY 460 TCCAAATTTTGGAACTGAGTTTCAGTAAGTACTGCCAGCTGGTCTCTCTCTCCCTTTT 519
 DB 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
 QY 520 GCTGAGATATCTGGAACAATTCCTCATAGATATAGATCTCTGCCATTGTGCTGGCT 579
 DB 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
 QY 580 GGTGTGCATGCAGAGTAGTGTCAACACCTTGGCGGCCAAATCAAT--- 627
 DB 182 LysProArgMetGluIleIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu 201
 QY 628 -----GTTCTAATTAGTAAGGTATTCC----- 651
 DB 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis 221
 QY 652 -----TATTATGAAGTCTTTGGCTAACAACGTCACATCTGTG 690
 DB 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSerSer 241
 QY 691 GTGGGACACTTATCTACAACTCTTTTACATTTAAGACAAGT----- 732
 DB 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
 QY 733 -----GGATGTAT 741
 DB 259 GlyPheSerAlaArgTyrTyrLeuValHisGlnGluProProGluAsnPheGlnCysAsn 278
 QY 742 GGAACACTGGGGTGGAGTGTGTGTCATCGCGGATCTCAAAATACAGCATCATCTGTG 801
 DB 279 AlaProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer--- 297
 QY 802 CTGGAGTGGACTGACCCACACAGGCGGCAAGAGAACAGTGGAAACCCCAAGCAGCGCTG 861
 DB 298 -----ThrPheSerAspGlyArgTyrTrpProGlnGlnSerArgLeu 311
 QY 862 AAAAAACCTGGACCGCTTGGCTGCTTTTCCCACTCATGAATACCACTGCTTACAATA 921
 DB 312 HisGlyAspAspAsnGlyTyrTrpProAsnValAspSerAsnLysGluTyrLeuGlnVal 331
 QY 922 GATTCTGAATAAGGAAAAAATAACAGGCATTATAACCACTGGA-----TCCACCATG 975
 DB 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
 QY 976 GTGGACACAATTTACTATGTCTGCTACAGAATCTCTACAGTATGATGGCAGAAA 1035
 DB 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
 QY 1036 TGACATGTGTACAGAGCCCTGGTGGCAGCAGATGAATATTTCAAGGAAACAAGAT 1095
 DB 372 TrpMetValTyrArgHisGly-----LysAsnHisLysValPheGlnAlaAsnAsp 389
 QY 1096 TATCACCAGGATGCGCTAATAACTTTTGGCCCAATTTATGCACCTTTTATTAGAGTG 1155
 DB 390 AlaThrGluLeuValLeuAsnLysLeuHisThrProLeuLeuThrArgPheIleArgIle 409
 QY 1156 AATCTTACCAATGGCAGCAGAAAATGGCATGAAATGGAGCTGCTCGGATGTCAGTTT 1215
 DB 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
 QY 1216 ATTCTAAAGGTGCTCTCCAAAACCTTACTCAACCTCCACCTCTCTCGGAAACAGCAATGAC 1275
 DB 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
 QY 1276 CTCAAAAACACTACAGCCCTCCAAAATAGCCNAAGTGTGTCGCCCAAAATTT----- 1329

Db 441 LeuSerGlyLeuLeuAlaAspThrGlnIleSerAlaSerSerThrArgGluThrLeuTrp 460
 QY 1330 -----ACCAACCACTACAACTCGCAGTACGAATTTCTCCACACAGAA 1380
 Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgAsnProGln 480
 QY 1381 CAACAACTGCCAGTCTGATGATCAGAAATACCTACGTAACCTCAATGTAACCAAGAT 1440
 Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
 QY 1441 GTA-----GCGCTGCTGCAGTCTTCTTCTCCCTGCTGCTG 1473
 Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaMetGluAlaArgAlaPhe 520
 QY 1474 GTCATGGTCTCACTACTCTCATTTCTATATTAGTGTGCTGGCAGCTGG----- 1524
 Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
 QY 1525 -----AGAAACAGAAAGAAAACCTGAAGCACC-----TATGACTTACTCTTACTGG 1572
 Db 541 ProArgThrGlnGlnProLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
 QY 1573 GACCGGCGAGTGGTGAAGAAATGAAGCAGTCTTCTTCTCCCAAGCAGTGGACCAT 1632
 Db 559 -----AspIleArgArgPhe----- 563
 QY 1633 GAGAAACCCAGTTCGCTATAGCAGCAGCAGGAATTAATCACCTGAGTCCACAGAGATC 1692
 Db 564 GluProValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
 QY 1693 ACCACAGTCTCGMGCTGACTCTGCAGAGTAT-----GCTCAGCAGCTGCTGAGGAGGA 1746
 Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
 QY 1747 ATTGTGGTACACTTCATCAAGATCTACTTTAAACCA-----GAAGAGGA 1794
 Db 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
 QY 1795 AAGAGAGCAGGCTAT-----GCAGACCTAGTACTCTTCACTCA 1833
 Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerGI 643
 QY 1834 CCAGGCGAGGAAGTTATCATGCTGCTGTAACCACTCCCAATAGGGCGCTGATAT 1893
 Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTyrMetTyrAspArgAlaLys 663
 QY 1894 GCAACCCCAATCATGATGACATGTGAGGCGACCCCA----- 1930
 Db 663 stpLeuGlnSerThrTrpIleSerAlaAsnProAsnAspArgThrPheProAspAs 683
 QY 1931 -----CAACTTCAG-----TTGGT 1944
 Db 683 pLysAsnPheLeuLysLeuGlnSerAspGlyArgGluGlyGlnPheGlyArgLeuIle 703
 QY 1945 CAGCCCTCCACATCCACTTCAAGGCTACGGGGAACCACTCCCACTAGTGGGAAC 2004
 Db 703 eSerProPro-ValHisLeu-----ProArgSerProValCysMetGluP 718
 QY 2005 TACAATACACTTCTCTC----- 2022
 Db 718 heGlnTyrGlnAlaMetMetGlyHisGlyValAlaLeuGlnValValArgGluAlaArg 738
 QY 2023 AGGACTGACAGCTGCTC-----CTGACCCAGCCGATGATACCCGAAAGCTGGG 2076
 Db 738 InGluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlySerGluTrpLysHisGly 757
 QY 2077 AAGCAGGCTTACTGCCCGCAGGATTTGGTGTGATGAGGTCGACAGCAGACACAGAA 2136
 Db 758 ArgIleIleLeuProSerTyrAsp---MetGluTyrGlnIle-----ValPheGluGly 774
 QY 2137 GTATCAGGACAGAGGATGGGATGTGATGTT 2172
 Db 775 ValIleGlyLysGlyArgSerGlyGluIleSerIle 786

RESULT 2
 NRP2_MOUSE
 ID NRP2_MOUSE STANDARD; PRT; 931 AA.
 AC O35373; O35373; O35374; O35376; O35377; O35378;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
 GN NRP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC STRAIN=BALB/C.
 RX MEDLINE=97470888; PubMed=9331348;
 RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
 RT "Neuropilin-2, a novel member of the neuropilin family, is a high
 affinity receptor for the semaphorins Sema E and Sema IV but not Sema
 III";
 RT Neuron 19:547-559(1997).
 RL
 CC -!- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
 AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
 CC -!- SUBUNIT: NEUROFILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
 NEUROFILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN E RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- ALTERNATIVE PRODUCTS: 6 ISOFORMS; A22 (SHOWN HERE), A0, A5, A17,
 B0 AND B5; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING CNS, PNS AND IN SOME
 NONNEURAL TISSUES INCLUDING LIMB BUDS, DEVELOPING BONES, MUSCLES,
 INTESTINAL EPITHELIUM, KIDNEY, LUNG AND SUBMANDIBULAR GLAND.
 CC -!- DEVELOPMENTAL STAGE: THE EXPRESSION PATTERN IS VERY DYNAMIC AND
 IS DEVELOPMENTALLY REGULATED.
 CC -!- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 NAM DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF022856; AAC53379.1; -
 CC EMBL; AF022854; AAC53377.1; -
 CC EMBL; AF022855; AAC53378.1; -
 CC EMBL; AF022857; AAC53380.1; -
 CC EMBL; AF022858; AAC53381.1; -
 CC EMBL; AF022861; AAC53382.1; -
 CC HSSP; P12259; 1C2T.
 CC MGD; MGI:1100492; Nrp2.
 CC InterPro; IPR000859; Cub_domain.
 CC InterPro; IPR000421; FA58_C.
 CC InterPro; IPR000998; NAM_domain.
 CC Pfam; PF00431; CUB; 2.
 CC Pfam; PF00629; NAM; 1.
 CC Pfam; PF00754; F5_F8_type_C; 2.
 CC SMART; SM00042; CUB; 2.
 CC SMART; SM00231; FA58C; 2.
 CC SMART; SM00137; NAM; 1.
 CC PROSITE; PS01180; CUB; 2.
 CC PROSITE; PS01285; FA58C_1; 2.
 CC PROSITE; PS01286; FA58C_2; 2.
 CC PROSITE; PS00060; NAM_2; 1.
 CC Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
 KW Alternative splicing.
 FT SIGNAL 1 20
 FT SIGNAL 1 20

Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerThrArgGluTyrLeuTyr 460
QY 1330 -----ACCAACCACTACCACTCCAGTAGCAATGAATTCCTCCACAGACAGAA 1380
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTyrPheProArgAsnProGln 480
QY 1381 CAAACAACCTGCCAGCTCCTGATATCAATACTACCGTACCTCCCAATGTAACCAAGAT 1440
Db 481 AlaGlnProGlyGluGluTyrLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1441 GTA-----GGCTGGCTGCAGTTCCTTCTTCTCCCTGGCTGG 1473
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGlnAlaArgAlaPhe 520
QY 1474 GTCATGGTCTCCTACTCTCATCTTCATATTAGTGTGCTGGCAGTGG----- 1524
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTyrGluTyrIleGlnAsp 540
QY 1525 -----AGAAACAGAAAGAAAGAACTGAAGGCACC-----TATGACTTACTCTACTGG 1572
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1573 GACCGGGCAGGTGGTGGAAAGATGAAGCAGTTCCTCTGCAAAAGCAGTGGACCAT 1632
Db 559 -----AspIleArgArgPheAspPro----- 565
QY 1633 GAGGAACCCAGTTCGCTATAGCAGCAGCAGGAAGTTAATCACTGAGTCCACAGAGAATC 1692
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTyrSerProAlaGlyIle 583
QY 1693 ACCACAGTCTGCGGCTGACTCTGCAGAGTAT-----GCTCAGCAGTGTGAGGAGGA 1746
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTyrThrAspSerLysProThrValGluThr 603
QY 1747 ATGCTGGTACACTTCATCAAGATCTACTTTAAACCA-----GAAGAAGGA 1794
Db 604 LeuGlyProThrLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
QY 1795 AAAGAAGCAGGTAT-----GCAGACCTAGTACTCTTCAACTCA 1833
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspAspLysAspLeuGln-LeuProSerGI 643
QY 1834 CCAGGCGCAGGAAGTTTATCATGCT-----ATGCTGACCACTCCCA 1875
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTyrValTyrAspHisAlaLys 663
QY 1876 A---TTACGGGCTCAGTATGCAACCCCAATCATCATGACATGTCAGGCGCACCCA-- 1930
Db 663 strLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 1931 -----CAACTTCAG----- 1939
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspLysArgArgGI 696
QY 1940 -----TTGGTCAGCCCTCCACATCCACTTTCAGGCTACGGGACACAA 1983
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 1984 CPTCCCTCCACTAGTGGGAATTAACATACATCTCTCTC----- 2021
Db 711 rSerProValCysMetGluPheGlnTyrGlnAlaMetGlyHisGlyValAlaLeuG 731
QY 2022 -----CAGACTGACAGCTGCTC-----CTCAGCCCGCCAGCCAG 2055
Db 731 lnValValArgGluAlaSerGlnGluSerLysLeuLeuTyrPheValIle-ArgGluAspGln 750
QY 2056 TATGATACCCGAAAGCTGGGACCGAGTCTTACCTGCCCGACAGCAATGGTGTACCAG 2115
Db 751 GlySerGluTyrLysHisGlyArgIleLeuLeuProSerTyrAsp---MetGluTyrGln 769
QY 2116 GTCCACAGACACACAGAGTATCAGGACGAGGAGGATCGGGAA 2163
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu 783

RESULT 3

ID NRPL_HUMAN STANDARD; PRT; 923 AA.
AC 014786; 060461;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuropilin-1 precursor (Vascular endothelial cell growth factor 165 receptor)
DE NRPI OR NRP OR VEGF165R.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM).
RX MEDLINE=97433084; PubMed=9288753;
RA He Z., Tessier-Lavigne M.;
RT "Neuropilin is a receptor for the axonal chemorepellent semaphorin III".
RL Cell 90:739-751(1997).
RN [2]
RP SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM), AND SEQUENCE OF 22-39.
RC TISSUE=Breast;
RX MEDLINE=98188099; PubMed=9529250;
RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor".
RL Cell 92:735-745(1998).
RN [3]
RP SEQUENCE FROM N.A. (SOLUBLE/SNRP1 ISOFORM), AND SEQUENCE OF 22-31.
RC TISSUE=Prostatic adenocarcinoma;
RX MEDLINE=20183929; PubMed=10688880;
RA Gagnon M.L., Bieleberg D.R., Gerechtman Z., Miao H.-Q., Takashima S., Soker S., Klagsbrun M.;
RT "Identification of a natural soluble neuropilin-1 that binds vascular endothelial growth factor: In vivo expression and antitumor activity".
RL Proc. Natl. Acad. Sci. U.S.A. 97:2573-2578(2000).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid form of vascular endothelial growth factor (VEGF) and of placenta growth factor-2, but only neuropilin-2 functions as a receptor for the 145-amino acid form of VEGF".
RL J. Biol. Chem. 275:18040-18045(2000).
CC -1- FUNCTION: THE MEMBRANE-BOUND ISOFORM IS A RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGF, THE VEGF-165 ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS. IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS.
CC -1- FUNCTION: THE SOLUBLE/SNRP1 ISOFORM BINDS VEGF-165 AND APPEARS TO INHIBIT ITS BINDING TO CELLS. IT MAY ALSO INDUCE APOPTOSIS BY SEQUESTERING VEGF-165. MAY BIND AS WELL VARIOUS MEMBERS OF THE SEMAPHORIN FAMILY. ITS EXPRESSION HAS AN AVERSE EFFECT ON BLOOD VESSEL NUMBER AND INTEGRITY.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. THE SOLUBLE/SNRP1 ISOFORM IS SECRETED.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; MEMBRANE-BOUND (SHOWN HERE) AND SOLUBLE/SNRP1; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: THE EXPRESSION OF MEMBRANE-BOUND AND SOLUBLE/SNRP1 ISOFORMS DOES NOT SEEM TO OVERLAP. THE MEMBRANE-BOUND ISOFORM IS EXPRESSED BY THE BLOOD VESSELS OF DIFFERENT TISSUES. IN THE DEVELOPING EMBRYO IT IS FOUND PREDOMINANTLY IN THE NERVOUS SYSTEM. IN ADULT TISSUES, IT IS HIGHLY EXPRESSED IN HEART AND PLACENTA; MODERATELY IN LUNG, LIVER, SKELETAL MUSCLE, KIDNEY

CC	AND PANCREAS; AND LOW IN ADULT BRAIN. THE SOLUBLE/SNRP1 ISOFORM IS
CC	FOUND IN LIVER HEPATOCYTES, KIDNEY DISTAL AND PROXIMAL TUBULES.
CC	-1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC	-1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC	-1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AF018956; AAC51759.1; -
DR	EMBL; AF016050; AAC12921.1; -
DR	EMBL; AF145712; AAF44344.1; -
DR	HSSP; P12259; ICZT.
DR	Genew; HGNC:8004; NRPI.
DR	MIM; 602069; -
DR	InterPro; IPR000859; CUB_domain.
DR	InterPro; IPR000421; FA58_C.
DR	InterPro; IPR001092; HLH_Basic.
DR	InterPro; IPR000998; MAM_domain.
DR	Pfam; PF00431; CUB; 2.
DR	Pfam; PF00629; MAM; 1.
DR	Pfam; PF00754; F5_F8_typeC; 2.
DR	PRINTS; PR00020; MAMDOMAIN.
DR	SMART; SM00042; CUB; 2.
DR	SMART; SM00231; FA58C; 2.
DR	SMART; SM00137; MAM; 1.
DR	PROSITE; PS01180; CUB; 2.
DR	PROSITE; PS01285; FA58C_1; 2.
DR	PROSITE; PS01286; FA58C_2; 2.
DR	PROSITE; PS00740; MAM_1; 1.
DR	PROSITE; PS50060; MAM_2; 1.
KW	Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW	Alternative splicing.
FT	SIGNAL 1 21
FT	CHAIN 22 923
FT	DOMAIN 22 856
FT	TRANSMEM 857 879
FT	DOMAIN 880 923
FT	DOMAIN 27 141
FT	DOMAIN 147 265
FT	DOMAIN 275 424
FT	DOMAIN 431 583
FT	DOMAIN 645 811
FT	DISULFID 27 54
FT	DISULFID 82 104
FT	DISULFID 147 173
FT	DISULFID 206 228
FT	DISULFID 275 424
FT	DISULFID 431 583
FT	CARBOHYD 150 150
FT	CARBOHYD 261 261
FT	CARBOHYD 300 300
FT	CARBOHYD 522 522
FT	CARBOHYD 842 842
FT	VARSPLIC 642 644
FT	VARSPLIC 645 923
FT	CONFLICT 26 26
FT	CONFLICT 749 749
FT	CONFLICT 855 855
SEQ	SEQUENCE 923 AA; 103120 MW; ADEADC4A849E5D57 CRC64;
Alignment Scores:	
Pred. No.:	2_04e-21
Score:	400.50
Percent Similarity:	42.33%
Best Local Similarity:	28.08%
Query Match:	10.09%
	Length: 923
	Matches: 130
	Conservative: 66
	Mismatches: 178
	Indels: 89

QY	55	GGAGCCCAAGC	AGAGTGATGGATGTGGACACACTGTACTAGGCCTCGAGACTGGAAACCCCTT	114
Db	21	ArgGlyGlnPro	AspProProCysGlyGlyArgLeuAsnSerLysAspAlaGlyTrile	40
QY	115	ACATCCATAAACTACCCACAGACCTATCCCAACAGCAGCTGTTTGTGAATGGGAGCCGT	174	
Db	41	ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluTrpIleValTyr	60	
QY	175	GTA---AAGTGGGAGAGAGAGTTCGCATCAAAATTGGT---	228	
Db	61	AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluTrileGluLys	80	
QY	229	TCTGATCTCTCACTTTAATTACTTCAGAAATTATAATGGAATTGGAGTCAGCAGAAGCT	288	
Db	81	Hisasp---CysLysTyrAspPheIleGluIleArgAspGlyAspSerGluSerAlaAsp	99	
QY	289	GAATAGGCAAACTACTGTGCTGGGTTGCAAAATGAAC-----	339	
Db	100	LeuLeuGlyLysHisCysGly-----AsnIleAlaProProHrIleIle	114	
QY	340	TCAAAAGCAATGAATACATCTCTTCATGAGTGGAAATCATGTTTCTGGACGCGGA	399	
Db	115	SerSerGlySerMetLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly	134	
QY	400	TTTTTGGCCTCATACTCTGTTATAGATAAACAAGATCTAATTACTTGTTTGGACACTGCA	459	
Db	135	PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer	150	
QY	460	TCCAATTTTTTGGAAACCTGAGTTCAGTAAGTACTGCGCAGCTGGTGTCTGCTCCCTTT	519	
Db	151	LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe	165	
QY	520	GCTGAGATATCTGAACAATTCCTCATGGATATAGAGATTCCTGCCATTTGTCATGGCT	579	
Db	166	ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla	181	
QY	580	GGTGTGATGCAGGAGTAGTGTCAACACAGCTTGGCGGCCAAATCAGT-----	627	
Db	182	LysProLysMetGluIleIleLeuGlnPheLeuIlePheAspLeuGluHisAspProLeu	201	
QY	628	-----GTTGTAATTAGTAAAGGTATTCC-----	651	
Db	202	GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis	221	
QY	652	-----TATTATGAAGTCTTTGGCTAACACCTCATCTCATCTGTG	690	
Db	222	ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerGluLeuArgSer	241	
QY	691	GTGGGACACTTATCTACAAGCTTTTATACA-----	720	
Db	242	ThrGlyIleLeuSerLeuThrPheHisThrAspMetAlaValAlaLysAspGlyPheSer	261	
QY	721	-----TTTATGACAAGTGGATGTGAAATCCCAAAAGCCAGGCTGAAAAACCT	750	
Db	262	AlaArgTyrTyrLeuValHisGlnGluProLeuGluAsnPheGlnCysAsnValProLeu	281	
QY	751	GGGATGAGTCTGGTGATCGCGGATCCCTCAATAACAGCATCATCTGCTGGAGTG	810	
Db	282	GlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSer-----	297	
QY	811	ACTGACCACACAGGGCAAGAGACAGTGTGAAACCCCAAAAGCCAGGCTGAAAAACCT	870	
Db	298	-----ThrTyrSerAspGlyArgTrpThrProGlnGlnSerArgLeuHisGlyAsp	314	
QY	871	GGACCGCTTGGCGTCTTTGGCCTGATGAAATACCATGGTGTACAAATAGATTGAAT	930	
Db	315	AspAsnGlyTrpThrProAsnLeuAspSerAsnLysGluTyrLeuGlnValAspLeuArg	334	
QY	931	AAGGAAAAAGAAATAACAGCGCATTAATACCACTGGA-----TCCACCATGGTGGAC	984	
Db	335	PheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGluThrGlnAsn	354	

[illegible]

QY	427	AACACAGACTAATTA	TCTGTTGTGCACACTGCATCCCAATT	TTTTTGGAACTCGAGTTCAGT	486
Db		:::	:::	:::	
QY	143	ArgGlyPro-----	-GlucySerGlnAsn	fyrThrAlaPro-----	154
Db		:::	:::	:::	
QY	487	AAGTACTGCCACGCTGGT	GCTGCTCCTTTTGCTGAGATA	TCTGGACAATTCCTCAT	546
Db		:::	:::	:::	
Db	155	-----	-ThrGlyValIleLeuLysSer	160	
QY	547	---GGATATAGAGATTCTCG	CATTGTCATG---	-GCTGGT	582
Db		:::	:::	:::	
Db	161	ProGlyPheProGluLys	TyrProAsnCysLeuGluCys	ThrTyrIleIlePheAlaPro	180
QY	583	GTCCATGCAGAGTAGTG-		-----TCAAACACG	609
Db		:::	:::	:::	
Db	181	LysMetSerGluIleIleLeu	GluPheGluSerPheAspLeu	GluGlnAspSerAsnPro	200
QY	610	TTGGCGGCCRAATC-		-----AGTGTGTAATTAGTAA	651
Db		:::	:::	:::	
Db	201	ProGlyGlyMetPheCysArg	TyrAspArgLeuGluIleTrp	AspGlyPheProGluVal	220
QY	652	-----TATTATGAAGTCTT	TGGCTAACACCTCACATCT	GTGGTG	693
Db		:::	:::	:::	
Db	221	GlyProHisIleGlyArg	TyrCysGlyGlnLysThrPro	GlyArgIleArgSerSer	240
QY	694	GCACACTTATCTACAGCT	TTTTTACA-		720
Db		:::	:::	:::	
Db	241	GlyValLeuSerMetValPhe	TyrThrAspSerAlaIleAlaLys	GluClyPheSerAla	260
QY	721	-----TTTAAGACAAGTGG	ATTATGGA	-----CysMetGlu	744
Db		:::	:::	:::	
Db	261	AsnTyrServalLeuGlnSer	SerIleSerGluAspPheLys		277
QY	745	ACACTGGGATGGAGCTGTG	TGTGTGATCGGATCCCTCAATA	TACACCATCTCTGTGTG	804
Db		:::	:::	:::	
Db	278	AlaLeuGlyMetGluSerGly	GluIleHisSerAspGluIleThrAla	SerSer-----	295
QY	805	GAGTGGAGTACCACAGCAG	GGCAGAAGACAGTGTGAAA	CCCCAAAAAGCCAGGCTGAAA	864
Db		:::	:::	:::	
Db	296	-----GlnTyrGlyThrAsn	TrpSerValGluArgSerArgLeu	Asn	309
QY	865	AAACTGGACGGCTTGGGCT	GCTTTTGGCACTGATGAATACCA	TGTTTCAAAATAGAT	924
Db		:::	:::	:::	
Db	310	TyrProGluAsnGlyIrr	ThrProGlyGluaspSerTyrLys	GluTrpIleGlnValasp	329
QY	925	TTGAATAAGGAAAGAAATA	ACAGGCATTATACCACTGGA-	-----TCCACCATGGTG	978
Db		:::	:::	:::	
Db	330	LeuglyLeuLeuArgPheVal	ThrAlaValGlyThrGlnGlyAla	IleSerLysGluThr	349
QY	979	GAGCACAAATTAATGTGCTG	CGCTACAGAACTCCTGACAGT	GATGATGGGCAGAAATGG	1038
Db		:::	:::	:::	
Db	350	LysLysLysTyrTyrVallys	ThrTyrArgValaspIleSerSer	AsnGlyGluaspTrp	369
QY	1039	ACTGTGTACAGAGAGCCTG	TGTGTGGAGCAAGATACAGATAT	TTCGAAGAAACAAGATAT	1098
Db		:::	:::	:::	
Db	370	IleSerLeuLysGluGly	-----AsnLysAlaIleIlePhe	GlnGlyAsnThrAsnPro	387
QY	1099	CACGAGATGTGCTAATACT	TTTTTGGCCACAATATTGGCACT	TTTATTAGATCAAT	1158
Db		:::	:::	:::	
Db	388	ThrAspValValLeuGly	ValPheSerLysProLeuIleThrArg	PheValArgIleLys	~407
QY	1159	CCTACCCAATGGCAGCAGAA	ANTGCCATGAATGGAGCTGCTCG	GATGTTCAGTTTATT	1218
Db		:::	:::	:::	
Db	408	ProValserTrpIleThrGly	IleSerMetArgPheGluValTyr	GlyCys-----	424
QY	1219	CCTAAAGGTGCTCTCCAA	AACTTACTCAACTCA	1254	
Db		:::	:::	:::	
Db	425	-----LysileThrAsp	TyrPro	430	
RESULT	6				
NRPL_RAT					
ID	NRPL_RAT	STANDARD;	PRT;	922 AA.	
Q90WJ9;					
DT	16-OCT-2001 (Rel. 40,	Created)			

Db	244	SerMetValPheTyrThrAspSerAlaIleAlaLysGluGlyPheSerAlaAsnTyrSer	263
Qy	721	-----TTTAAAGACAAGTGGATGTTATGGACAACACTGGGG	753
Db	264	ValLeuGlnSerSerIleSerGluAspPheLys-----CysMetGluAlaLeuGly	280
Qy	754	ATGGAGTCTGGTGTATCGCGGATCCTCAAAATAACAGCATCATCTGTGCTGGAGTGGACT	813
Db	281	MetGluSerGlyGluIleHisSerAspGlnIleThrAlaSerSer-----	295
Qy	814	GACCACACAGGCAAGAGAACACTTGGAAACCCAAAAAGCCAGGCTGAAAAAACCTGGA	873
Db	296	-----GlnTyrGlyThrAsnTrpSerValGluArgSerArgLeuAsnTyrProGlu	312
Qy	874	CCGCTTGGCTGCTTTGCCACTGATGAATACCAGTGGTTACAAATAGATTGTAATAAG	933
Db	313	AsnGlyTrpThrProGlyGluAspSerTyrArgGluTrpIleGlnValAspLeuGlyLeu	332
Qy	934	GAAGAAGAAATAACAGGCATTATACCACTGGA-----TCCACCATGGTCGAGCACAA	987
Db	333	LeuArgPheValThrAlaValGlyThrGlnGlyAlaIleSerLysGluThrLysLysLys	352
Qy	988	TACTATGTGTCGCTACAGAACTCCTGACAGTGTATGGGACAGAAATGACGTGTGAC	1047
Db	353	TyrrTyrValLysThrTyrArgValAspIleSerSerAsnGlyGluAspTrpIleThrLeu	372
Qy	1048	AGAGAGCGTGTGTGGAGCAAGATAAGATATTTCACGAGAAACAAAGATATTATCACCAG	1107
Db	373	LysGluGly-----AsnLysAlaIleIlePheGlnGlyAsnThrAsnProThrAspVal	390
Qy	1108	GTGGCTTAATACTTTTGGCCACCAATATTGACAGCTTTTATTAGATGAATCCTACCCAA	1167
Db	391	ValPheGlyValPheProLysProLeuIleThrArgPheValArgIleLysProAlaSer	410
Qy	1168	TGGCAGCAGAAAATGGCATGAAATGGAGCTGCTCGGATGTCAGTTATTCTCTAAAGGT	1227
Db	411	TrpGluThrGlyIleSerMetArgPheGluValTyrGlyCys-----	424
Qy	1228	CGTCTCCAAAACCTACTCAACCTCCA	1254
Db	425	-----LysIleThrAspTyrPro	430
RESULT 7			
NRPI_CHK			
ID	NRPI_CHK	STANDARD;	PRT; 914 AA.
AC	P79795;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Neuropilin-1 precursor (A5 protein).		
GN	NRPI OR NRP.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauiria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
KC	STRAIN-White leghorn; TISSUE-Embryonic brain;		
RX	MEDLINE=95324761; PubMed=7601310;		
RA	Takagi S., Kasuya Y., Shimizu M., Matsura T., Tsuboi M., Kawakami A.,		
RA	Fujisawa H.;		
RT	"Expression of a cell adhesion molecule, neuropilin, in the		
RT	developing chick nervous system."		
RL	Dev. Biol. 170:207-222(1995).		
CC	-!- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE		
CC	CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF		
CC	CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS, IN THE NERVOUS		
CC	SYSTEM. IT MEDIATES THE CHEMOPULSANT ACTIVITY OF SEMAPHORINS (BY		
CC	SIMILARITY). SEEMS TO HAVE CALCIUM-INDEPENDENT CELL ADHESION		
CC	PROPERTIES.		
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-!- TISSUE SPECIFICITY: DEVELOPING NERVOUS SYSTEM; OPTIC TECTUM		

Db		59	ProGluProTyrGlnArgIleMetIleAsnPheAsnProHisPheaspLeuGluasArg	78
Qy		232	GATTCTGTCTACTTTAAATTAATTCAGAAATTAATAATGCAATTTGGAGTCAGCAGAAC TGAA	291
			: ::: :: :	
Db		79	Asp---CysLysTyrAspTyrValcLuValileAspGlyAspAsnAlaGluGlyArgLeu	97
Qy		292	ATAGCGAANAATCTGTGTCTGGGTTGC AATGAACCATTCAATTAATGATCAAAGGCAAT	351
			: :	
Db		98	TripGlyLysTyrCys-----GlyLysIleAlaProProLeuValSerSerGlyPro	115
Qy		352	GAATCATCATTTGCTTCATGAGTGGAATCCATCTTTCTTGACGCCGATTTTTGGCCCTCA	411
			::: :: : :	
Db		116	TyrLeuPheIleLysPheValSerAspTyrGluThrHISgLyAlaGlyPheSerIleArg	135
Ov		412	TACTCTGTTATAGATAAAA-----	499

D	b		136	TyrGluValPheLysArgGlyProGluCysSerArgAsnPheThrSerSerSerGlyMet	1355
				::::	
Q	y		430	-----CAAGATCTAATTACTTGT-----	447
				::: :::	
D	b		156	IleLysSerProGlyPheProGluLysTyProAsnSerLeuGluCysThrTyIleIle	175
Q	y		448	-----TTGGACACTGCATCCAACTTTTGGAACCTCGAG	480
				:	
D	b		176	PheAlaProLysMetSerGluIleIleLeuGluPheGluSerPheGluLeuGluProASP	195
Q	y		481	TTCAGTAGACTGCCAGCTGGTGTCCTCTTCTGCTGAGATATCTCGAACAAAT	540
		:::		::: :::	
D	b		196	SerAsn-----ThrProGlyGlyAlaPheCysArgTyAspArgLeuGlu----	211
Q	y		541	CCTCATGGATAGAGATTCCTCGCCA-----TTGTGCATGGCTGGTGTGCAT	588
		:		:::	
D	b		212	TrpaspGlyPheProaspValglyProHisIleGlyArgTyCysGlyGlnAsnAsnPro	231

[illegible][illegible][illegible]

FT	DOMAIN	348	683	F5/8 TYPE A 2.	FT	MOD_RES	1631	1631	SULFATION (POTENTIAL).
FT	DOMAIN	348	525	PLASTOCYANIN-LIKE 3.	FT	CARBOHYD	153	153	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DOMAIN	535	683	PLASTOCYANIN-LIKE 4.	FT	CARBOHYD	225	225	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DOMAIN	691	1611	B.	FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DOMAIN	1168	1539	41 X 9 AA APPROXIMATE TANDEM REPEATS OF T-L-S-P-D-L-[GS]-[RQ]-T.	FT	CARBOHYD	297	297	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1168	1176	1.	FT	CARBOHYD	382	382	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1177	1185	2.	FT	CARBOHYD	460	460	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1185	1193	3.	FT	CARBOHYD	467	467	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1195	1203	4.	FT	CARBOHYD	553	553	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1204	1212	5.	FT	CARBOHYD	741	741	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1213	1221	6.	FT	CARBOHYD	752	752	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1222	1230	7.	FT	CARBOHYD	760	760	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1231	1239	8.	FT	CARBOHYD	776	776	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1240	1248	9.	FT	CARBOHYD	782	782	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1249	1257	10.	FT	CARBOHYD	899	899	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1258	1266	11.	FT	CARBOHYD	960	960	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1267	1275	12.	FT	CARBOHYD	1048	1048	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1276	1284	13.	FT	CARBOHYD	1057	1057	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1285	1293	14.	FT	CARBOHYD	1066	1066	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1294	1302	15.	FT	CARBOHYD	1073	1073	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1303	1311	16.	FT	CARBOHYD	1089	1089	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1312	1320	17.	FT	CARBOHYD	1174	1174	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1321	1329	18.	FT	CARBOHYD	1480	1480	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1330	1338	19.	FT	CARBOHYD	1537	1537	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1339	1347	20.	FT	CARBOHYD	1597	1597	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1348	1356	21.	FT	CARBOHYD	1737	1737	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1357	1365	22.	FT	CARBOHYD	1886	1886	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1366	1374	23.	FT	CARBOHYD	2044	2044	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1375	1383	24.	FT	CARBOHYD	2243	2243	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1384	1392	25.	SQ	SEQUENCE	2258	2258	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	REPEAT	1393	1401	26.	US-10-060-830-2 (1-2190) x FA5_PIG (1-2258)				
FT	REPEAT	1402	1410	27.	Qy 724 AAGACAAGTGGATGTTATGGAACTGGGATGGAGTCTGGTGTGATCGCGGATCCTCAA 783				
FT	REPEAT	1411	1419	28.	Db 2096 GluValAsnGlySerThrProLeuGlyMetGluSerGlyAsnIleLysAsnGluIn 2115				
FT	REPEAT	1420	1428	29.	Qy 784 ATAACAGCATCATCTGTCTGGAG-----TGGACTGACACACAGGCGCAAGAGAACAGT 837				
FT	REPEAT	1429	1437	30.	Db 2116 IleThrAlaSerSerPheLysLysSerTrpTrpGlyAspTyr----- 2129				
FT	REPEAT	1438	1446	31.	Qy 838 TGGAAACCCAAAGAGCAGGCTGAAAGAACCTGGA-----CCGCTTGGCTGCTTTT 891				
FT	REPEAT	1447	1455	32.	Db 2130 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2149				
FT	REPEAT	1456	1464	33.	Qy 892 GCCACTGATGATACCACTGGTTACAAATAGATTGAATAAGGAAAGAAATAACAGCG 951				
FT	REPEAT	1465	1473	34.	Db 2150 AlaAsnAsnAsnGlnTrpLeuGlnIleAspLeuLysIleLysIleThrAla 2169				
FT	REPEAT	1474	1482	35.	Qy 952 ATTATAACCACTGGATCCACCATGGTGGAGCAACAATTACTATGTGCTGCTTACAGAAATC 1011				
FT	REPEAT	1483	1491	36.	Db 2170 IleThrThrGlnGlyCysLysSerLeuSerSerGluMetTyrValLysArgTyrThrIle 2189				
FT	REPEAT	1492	1500	37.	Qy 1012 CTGTACAGTGTATGGGCGAGAAATGGAGTGTGTACAGAGACCTGGTGTGGAGCAAGAT 1071				
FT	REPEAT	1501	1509	38.	Db 2190 GlnTyrSerAspArgGlyValGlnTrpLysSerTyrArgLysSerSerMetValAsp 2209				
FT	REPEAT	1510	1518	39.	Qy 1072 AAGATATTCAAGGAACAAAGATATATACACAGGATGTCGTAATAACATTTTGGCACC 1131				
FT	REPEAT	1519	1527	40.	Db 2210 LysIlePheGluGlyAsnAsnIleLysGlyHisValLysAsnPheAsnProPro 2229				
FT	REPEAT	1531	1539	41.	Qy 1132 ATTATTGCAGTATTATTAGTAGTGAATCCTACCAATGTAATGAGAGAGAGAGAGAGAG 1191				
FT	DOMAIN	1616	1785	F5/8 TYPE A 3.	Db 2230 IleIleSerArgPheIleArgIleIleProLysMetTrpAsnGlnSerIleAlaLeuArg 2249				
FT	DOMAIN	1616	1785	PLASTOCYANIN-LIKE 5.					
FT	DOMAIN	1795	1941	PLASTOCYANIN-LIKE 6.					
FT	DOMAIN	1942	2095	F5/8 TYPE C 1.					
FT	DOMAIN	2100	2255	F5/8 TYPE C 2.					
FT	SITE	737	738	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).					
FT	SITE	1029	1030	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).					
FT	SITE	1611	1612	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).					
FT	DISULFID	167	193	PROBABLE.					
FT	DISULFID	499	525	PROBABLE.					
FT	DISULFID	1759	1785	PROBABLE.					
FT	DISULFID	1941	2095	BY SIMILARITY.					
FT	DISULFID	2100	2255	BY SIMILARITY.					
FT	MOD_RES	692	692	SULFATION (POTENTIAL).					
FT	MOD_RES	696	696	SULFATION (POTENTIAL).					
FT	MOD_RES	724	724	SULFATION (POTENTIAL).					
FT	MOD_RES	726	726	SULFATION (POTENTIAL).					
FT	MOD_RES	745	745	SULFATION (POTENTIAL).					
FT	MOD_RES	1560	1560	SULFATION (POTENTIAL).					
FT	MOD_RES	1576	1576	SULFATION (POTENTIAL).					
FT	MOD_RES	1581	1581	SULFATION (POTENTIAL).					
FT	MOD_RES	1584	1584	SULFATION (POTENTIAL).					
FT	MOD_RES	1588	1588	SULFATION (POTENTIAL).					

Alignment Scores:

Pred. NO.:	1.5e-18	Length:	2258
Score:	362.00	Matches:	79
Percent Similarity:	58.43%	Conservative:	18
Best Local Similarity:	47.59%	Mismatches:	59
Query Match:	9.12%	Indels:	10
DB:	1	Gaps:	3


```
FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1094 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1490 1490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1690 1690 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1839 1839 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1997 1997 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2196 2196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 587 592 NGLTPA -> T (IN VARIANT 2).
SQ SEQUENCE 2211 AA; 248981 MW; CBBF90B738667C45 CRC64;

Alignment Scores:
Pred. No.: 1,76e-18 Length: 2211
Score: 361.00 Matches: 79
Percent Similarity: 57.83% Conservative: 17
Best Local Similarity: 47.59% Mismatches: 60
Query Match: 9.10% Indels: 10
DB: 1 Gaps: 3

US-10-060-830-2 (1-2190) x FA5_BOVIN (1-2211)
QY 724 AAGACAGTGGATGTTATGGAACACTGGGAGTGGTGTGATCGGATCCTCAA 783
Db 2049 GluValAsnGlyCysSerThrProLeuGlyMetGluSerGlyLysIleGluAsnLysGln 2068
QY 784 ATAACAGCATCATGTGTCTGGAG-----TGGACTGACCACACAGCGGCAAGACAGT 837
Db 2069 IleThrAlaSerSerPheLysLysSerTrpTTP-----GlyAsnTyr 2082
QY 838 TGAACACCCAAACAGCCAGGCTGAAAAACCTGGA-----CCGCTTGGGCTGCTTTT 891
Db 2083 TrpGluProPheLeuAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2102
QY 892 GCACATGATGAATACACAGTGGTTACAAATAGATTTCGAATAAGAAAGAAATAACAGGC 951
Db 2103 AlaAsnAsnAsnGlnTrpLeuGlnIleAspLeuLysIleLysLysIleThrAla 2122
QY 952 ATTATAACCATCGATCCACCATCGTGGAGCACAAATTAATGCTGCTGCCTACAGAATC 1011
Db 2123 IleValThrGlnGlyCysLysSerLeuSerGluMetTyrValLysSerTyrThrIle 2142
QY 1012 CTTGACATGATGATGGCAGAAATGAGTGTGTACAGAGAGCCTGGTGGAGCAAGAT 1071
Db 2143 HisTyrSerAspGlnGlyThrAspTTPTrpLysProTyrArgGluLysSerSerMetValAsp 2162
QY 1072 AAGATATTTCAAGGAACAAAGATTATCACCAGGATGCGTAATAACTTTTGCACCA 1131
Db 2163 LysIlePheGluGlyAsnAsnValArgGlyHisValLysAsnPheAsnProPro 2182
QY 1132 ATTATTGACAGTGTATTAGAGTGAATCCTACCAATGCGAGCAGCAAAATGTCATGAA 1191
Db 2183 IleIleSerArgPheIleArgIleIleProLysThrTrpAsnGlnSerIleAlaLeuArg 2202
QY 1192 ATGAGCTGCTCGGATGT 1209
Db 2203 LeuGluLeuPheGlyCys 2208

RESULT 11
MFGM_BOVIN
ID MFGM_BOVIN STANDARD; PRT; 427 AA.
AC Q95114; Q27959; P79344;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE (MGP57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE SP47) (BP47) (Components 15/16).
GN MFGEM.
```

```
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC STRAIN-Holstein; TISSUE-Mammary gland;
RA MEDLINE=97008954; PubMed=8856064;
RX Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT milk fat globules.";
RL Eur. J. Biochem. 240:628-636(1996).
RN [2]
RP SEQUENCE OF 18-427 FROM N.A.
RC TISSUE-Mammary gland;
RX MEDLINE=96125736; PubMed=8541316;
RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA Matsuda T.;
RT "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
RT monoclonal antibodies raised against bovine milk fat globule
RT membrane.";
RL Biochim. Biophys. Acta 1245:385-391(1995).
RN [3]
RP SEQUENCE OF 19-427 FROM N.A.
RC TISSUE-Testis;
RA Ensslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC TISSUE-Milk;
RA MEDLINE=93250576; PubMed=8485470;
RX Mather I.H., Banghart L.R., Lane W.S.;
RT "The major fat-globule membrane proteins, bovine components 15/16 and
RT guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
RT containing epidermal growth factor-like and factor V/VIII-like
RT sequences.";
RL Biochem. Mol. Biol. Int. 29:545-554(1993).
CC -1- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
CC LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN.
CC -1- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
CC -1- PTM: THE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH
CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X91895; CAA62997.1; -.
CC EMBL; S80643; AAB35894.2; -.
CC EMBL; Y11719; CAA72406.1; -.
CC HSSP; P00740; 1IXA.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR000421; FA58_C.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00231; FA58C; 2.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
```


Qy 724 AAGACAAGTGGATGTTATGAACACTGGGATGGAGTGTGTGTGTGATCGCGGATCCCAA 783
 Db 2062 GluValAsnGlyCysSerThrProLeuGlyMetGluAsnGlyLysIleGluAsnLysGln 2081
 Qy 784 ATACACATCATCTGTCTGGAG-----TGGACTGACACACAGGCGAAGAACAGT 837
 Db 2082 IleThrAlaSerSerPheLysLysSerTrpTrpGlyAspTyr----- 2095
 Qy 838 TGGAAACCCAAAGCCAGCTGAAACCTGGA-----CGCCTTGGCTGCTTTT 891
 Db 2096 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2115
 Qy 892 GCCACTGATGAATACCACTGGTGTACAAATAGATTGAAGAAAGAAATAACAGGC 951
 Db 2116 AlaAsnAsnLysGlnTrpLeuGluIleAspLeuLysLysIleThrAla 2135
 Qy 952 ATTATACCACTGATCCACATGGTGGACACAATTACTATGTGTCTGCTACAGAAATC 1011
 Db 2136 IleIleThrGlnGlyCysLysSerLeuSerSerGluMetTyrValLysSerTyrThrIle 2155
 Qy 1012 CTGTACAGTGTATGGCGAGAAATGCTGTGTACAGAGAGCTGTGTGGACAGAT 1071
 Db 2156 HisTyrSerGluGlnGlyValGluTrpLysProTyrArgLeuLysSerSerMetValAsp 2175
 Qy 1072 AAGATATTCAAGAAACAAAGATTATCACCAGGATGTGGTATTAATTAATTAATTAAT 1131
 Db 2176 LysIlePheGluGlyAsnThrAsnThrLysGlyHisValLysAsnPhePheAsnProPro 2195
 Qy 1132 ATTATTCAGCTTTTATTAGATGAATCCATACCAATGCGCAGAGAAATATGCGATGAA 1191
 Db 2196 IleIleSerArgPheIleArgValIleProLysThrTrpAsnGlnSerIleThrLeuArg 2215
 Qy 1192 ATGGAGCTGCTCGGATGT 1209
 Db 2216 LeuGluLeuPheGlyCys 2221

RESULT 13
 MFGM_HUMAN STANDARD; PRT; 387 AA.
 AC 008431;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)
 DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
 GN MFG-E8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Breast, and Breast carcinoma;
 RX MEDLINE=96213908; PubMed=8639264;
 RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.,
 RT "Cloning and sequence analysis of human breast epithelial antigen
 BA46 reveals an RGD cell adhesion sequence presented on an epidermal
 growth factor-like domain.";
 RL DNA Cell Biol. 15:281-286(1996).
 RN [2]
 RP SEQUENCE OF 170-387 FROM N.A.
 RC TISSUE-Mammary gland;
 RX MEDLINE=91371351; PubMed=1909932;
 RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
 RT Ceriani R.L.;
 RT "A Mr 46,000 human milk fat globule protein that is highly expressed
 in human breast tumors contains factor VIII-like domains.";
 RL Cancer Res. 51:4994-4998(1991).
 RN [3]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RC TISSUE-Milk;
 RX MEDLINE=98194924; PubMed=9535276;
 RA Gluffrida M.G., Cavaletto M., Giunta C., Conti A.,

Godovac-Zimmermann J.;
 "Isolation and characterization of full and truncated forms of human
 breast carcinoma protein BA46 from human milk fat globule membranes.";
 J. Protein Chem. 17:143-148(1998).
 RN [4]
 RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
 RX MEDLINE=99342076; PubMed=10411933;
 RA Haegqvist B., Naeslund J., Sletten K., Westermarck G.T., Mucchiano G.,
 Tjernberg L.O., Nordstedt C., Engstrom U., Westermarck P.;
 "Medin: an integral fragment of aortic smooth muscle cell-produced
 lactadherin forms the most common human amyloid.";
 Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=97405885; PubMed=9260929;
 RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
 "Lactadherin (formerly BA46), a membrane-associated glycoprotein
 expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp.
 (RGD)-dependent cell adhesion.";
 DNA Cell Biol. 16:861-869(1997).
 CC -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
 CC -!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
 MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
 CC -!- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
 AMINO ACID 264 AND 273.
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U58516; AAC50549.1; -;
 CC EMBL: S56151; AAB19771.1; -;
 CC HSSP: P08709; LBF9.
 CC Genew: HGNC:7036; MFG-E8.
 CC MIM: 602281; -;
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000421; FA58_C.
 CC Pfam: PF00008; EGF; 1.
 CC Pfam: PF00754; F5_F8_type_C; 2.
 CC SMART: SM00181; EGF; 1.
 CC SMART: SM00231; FA58C; 2.
 CC PROSITE: PS00022; EGF_1; 1.
 CC PROSITE: PS01186; EGF_2; 1.
 CC PROSITE: PS01285; FA58C_1; 2.
 CC PROSITE: PS01286; FA58C_2; 2.
 KW Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
 FT SIGNAL 1 23
 FT CHAIN 24 387 LACTADHERIN.
 FT CHAIN 202 387 LACTADHERIN; SHORT FORM.
 FT CHAIN 268 317 MEDIN.
 FT DOMAIN 24 67 EGF-LIKE.
 FT DOMAIN 70 225 F5/8 TYPE C 1.
 FT DOMAIN 230 387 F5/8 TYPE C 2.
 FT SITE 27 38 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 26 38 BY SIMILARITY.
 FT DISULFID 32 55 BY SIMILARITY.
 FT DISULFID 57 66 BY SIMILARITY.
 FT DISULFID 70 225 BY SIMILARITY.
 FT DISULFID 212 216 BY SIMILARITY.
 FT DISULFID 230 387 BY SIMILARITY.
 FT CARBOHYD 238 238 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 325 325 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC...) (POTENTIAL).


```
QY 517 TTGCTGAGATATCTGGAACAATTCCTCAT-----GGATATAGAGATTCTCGCCA 567
    |||||
Db 217 -----SerGlyThrGlyAspLysGluPheMetGlyAsnGlnAspAsnAsnSer 232
    |||||
QY 568 TTGTCATGGCTGTCATGCAGGAGTATGTCACACAGTTCGGCGGCCAAATCAGT 627
    |||||
Db 233 Leu-----LysIleasnMetPheAsnProThrLeuGluAlaGlnTyr--- 246
    |||||
QY 628 GTTGTAATTAGTAAGGTATTCCCTATTGAAAGTCTCTTGGCTAACAAACGTCACATCT 687
    |||||
Db 247 -----IleArgLeuTyrProValSerCysHisArgGlyCysThr--- 259
    |||||
QY 688 GTGGTGGACACTATCTACAGTCTTTTACATTATAGACAAGTGGATGTTATGGAACA 747
    |||||
Db 260 -----LeuArgPheGluLeuGlyCysGluLeuHisGlyCysSerGluPro 275
    |||||
QY 748 CTGGGATGGAGTCTGGTGTGTCGGGATCTCTCAATAACAGCATCATCTGTGCTGGAG 807
    |||||
Db 276 LeuGlyLeuLysAsnAsnThrIleProAspSerGlnIleThrAlaSer-----Ser 293
    |||||
QY 808 TGGACTGACCACAGGACAGAGAACAGTTGGAAACCCCAAAAGCCAGGCTGAAAAAA 867
    |||||
Db 294 TyrLysThrTrpAsnLeuArgAlaPheGlyTrpTyrProHisLeuGlyArgLeuAspAsn 313
    |||||
QY 868 CTTGGACCG-----CCTTGGCTGCTTTTGGCCACTGATGAATACAGTGGTTACAAATA 921
    |||||
Db 314 GlnGlyLysIleAsnAlaTrpThrAlaGlnSerAsnSerAlaLysGluTrpLeuGlnVal 333
    |||||
QY 922 GATTTGAATAAGGAAGAAATAACAGCATATATACCACTGGATCCACCATTGGTGAG 981
    |||||
Db 334 AspLeuGlyThrGlnLysLysValThrGlyIleIleThrGlnGlyAlaArgAspPheGly 353
    |||||
QY 982 CACAATTACTATGTCCTGCCTACAGATCCCTGTACAGTGTATGCGGAGAAATGGACT 1041
    |||||
Db 354 HisIleGlnTyrValAlaSerTyrLysValAlaHisSerAspGlyValGlnTrpThr 373
    |||||
QY 1042 GTGTACAGAGACCTGGTGTGGAGCAAGATAAGATATTTCAAGGAACAAAGATTATCAC 1101
    |||||
Db 374 ValTyrGluGluGlnGlyThr-----SerLysValPheGlnGlyAsnLeuAspAsnAsn 391
    |||||
QY 1102 CAGGATGTGCGTAATAACTTTTGGCCCAATATTGCACGTTTATTAGATGATCCT 1161
    |||||
Db 392 SerHisLysLysAsnIlePheGluLysProPheMetAlaArgTyrValArgValLeuPro 411
    |||||
QY 1162 ACCCAATGGCAGCAAAATTCCTGAAATCGAGCTGCTCGGATGT 1209
    |||||
Db 412 LeuSerTrpHisAsnArgIleThrLeuArgLeuGluLeuGlyCys 427
    |||||
```

Search completed: January 21, 2003, 09:37:43
Job time : 37.3647 secs


```

DR      ENBL: AF387547: AAL30178.1; -.
DR      InterPro: IPR000859; CUB_domain.
DR      InterPro: IPR004211; FAS3_C.
DR      InterPro: IPR004043; LCCL_dom.
DR      Pfam: PF00431; CUB; 1.
DR      Pfam: PF00754; F5_F8_type_C; 1.
DR      ProSite: PS01180; CUB; 1.
DR      DNA-binding: Transcription regulation.
DR      SK  DNA-binding: 775 AA; 85034 MW; 3D06F81EF2337010 CRG64;
DR      QW  SEQUENCE

```

Alignment Scores:

Pred. No.:	0	Length:	775
Score:	388.00	Matches:	729
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.96%	Indels:	0
DB:	4	Gaps:	0

US-10-060-830-2 (1-2190) x Q96PD2 (1-775)

QY	1	ATGCCCTGTTCTCCTCGTCTTACTTGTCCTCTGCTGCTCAGGACGCCTGGAGCC	60
Dbb	47	MetProLeuPheLeuLeuLeuValLeuLeuLeuLeuLeuAspAlaGlyAla	66
QY	61	CAGCAAGGTGATGGATGTGCACACACTGTACTAGGCCCTCAGAGTGGAAACCCTTACATCC	120
Dbb	67	GlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSer	86
QY	121	ATAAATACCACAGAGACTATCCACAAGCACACTGTTGTGAATGGAGAGATCCGTGTAAG	180
Dbb	87	IleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluLeuArgValLys	106
QY	181	ATGGGAGAGAGTCCGATCAAAATTTGGTGACTTTGACATTGAAGATCTTGATCTTGT	240
Dbb	107	MetGlyGluArgValArgIleLysPheGlyAspPheAspileGluaspSeraspSerCys	126
QY	241	CACPTTAATTACTTGACAATTTATATGGAATTTGGAGTCCAGCAGAAGCTGAAATAGCAAA	300
Dbb	127	HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys	146
QY	301	TACTGTGCTGGGGTTGCAAAATGAACCATTCAAATGAATCAAAGGCCAATGAAATACAA	360
Dbb	147	TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr	166
QY	361	TTGCTGTTCAATGAGTGAATCCATGTTTTCPNGACGGGATTTTGGCCCTCATACTCTGT	420
Dbb	167	LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal	186
QY	421	ATAGATAAACAGATCAATTAATCTGTTGGACACTGCATCCAAATTTTGGNACCTGAG	480
Dbb	187	IleaspLysGlnaspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu	206
QY	481	TTCAGTAAGTACTGCCAGCTGGTGTCTGCTCTCTTTGCTGAGATATCTGGAAACAAT	540
Dbb	207	PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluLeuSerGlyThrIle	226
QY	541	CCTCATGGATATAGAGATTCTCCGCCATTGTGCATGGCTGGTGCATGCAGAGTAGTG	600
Dbb	227	ProHisGlyTyrArgaspSerSerProLeuCysMetalaglyValHisAlaGlyValVal	246
QY	601	TCAAACAGTTGGGGGCCCAAAATCAGTGTGTGAATTAGTAAGGTATCCCTATTATGAA	660
Dbb	247	SerAsnThrLeuGlyGlyGlnIleSerValIleSerLysGlyIleProTyrTyrGlu	266
QY	661	AGTTCCTTTGGCTACACAGTCACATCTGTGGTGGGACACTTATCTACAAGTCTTTTACA	720
Dbb	267	SerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr	286
QY	721	TTTAAGACAAGTGGATGTTATGAAACACTGGGGATCGAGTCTGGTGTGATCCGGATCT	780
Dbb	287	PheLysThrSerGlyCysTyrGlyThrLeuGlyMecGluSerGlyValIleAlaAspPro	306
QY	781	CAATAACACAGCATCATCTGTCGTGGAGTGGACTGACCACACAGGGGCAAGAACACTTGG	840

Db	307	GlnIleThrAlaSerValLeuGlnIuTrpThrAspHisThrGlyGlnIuAsnSerTrp	326
Qy	841	AAACCCAAAAGCCAGGCTGAAAAACCTGGACGCCCTGGCGTCTTTGGCCACTGAT	900
Db	327	LysProLysLysAlaArgLeuLysLysProGlyProProTrpAlaAlaAlaHealThrAsp	346
Qy	901	GAATACCAGTGGTTACAAATAGATTGCAATAGGAAAAGAAAATACAGGCATTATACC	960
Db	347	GlnTyrglnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThr	366
Qy	961	ACTGGATCCACCATGGTGGAGCACAAATTACTATGTGTCTGCGCTACAGAAATCCTGTACAGT	1020
Db	367	ThrGlySerThrMetValGluHisAsnTyrtyrValSerAlaItyrArgIleLeuTySer	386
Qy	1021	GATCATGGGCAGAAATGGACTGTGTACAGAGAGCCGTGGTGGACCAAGATAAGATATTT	1080
Db	387	AspAspGlyGlnLysTrpThrValTyraArgGluProGlyValGluGlnAspLysIlePhe	406
Qy	1081	CAAGGAAACAAAGATTATCACCAAGGATGCGCGTAATAACTTTTTGCCCACTAATTATGCA	1140
Db	407	GlnGlyAsnLysAspTyraHisGlnAspValArgAsnAsnPheLeuProIleIleAla	426
Qy	1141	CGTTTTATTAGTGAATCTCTACCCAAATGGCAGCAGAAAATTCGCAATGGAGCTG	1200
Db	427	ArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeu	446
Qy	1201	CTCGGATGCAGTTATTTCCTAAAGTCGTCTCCAAAACTTACTCACTCCACCTCCT	1260
Db	447	LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProPro	466
Qy	1261	CGGAACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAATACGCAAGAGTCGTGCC	1320
Db	467	ArgAsnSerAsnAspLeuLysAsnThrThrAlaProLysIleAlaLysLysArgAla	486
Qy	1321	CCAAAATTACGCAACACTACAACTCGGAGTAGCAATGAAATTTCTCTGCACAGACAGAA	1380
Db	487	ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu	506
Qy	1381	CAAACTACTGCCAGTCTGATATCAGAAATACTACCGTAACCTCAAATGTAAACCAAGAT	1440
Db	507	GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp	526
Qy	1441	GTAGCGTGGCTGCAGTCTTGTCCTCGTGGTATGGTCTCTCACTACTCTCATTTCTC	1500
Db	527	ValAlaLeuAlaAlaValLeuValProValLeuValIleMetValLeuThrThrLeuIleLeu	546
Qy	1501	ATATTAGTGTGCTGGCACTGGAGAACAGAAAGAAAAAACHTGAAGCACCTATGAC	1560
Db	547	IleLeuValCysAlaItrpHisItrpArgAsnArgLysLysLysThrGluGlyThrTyraSp	566
Qy	1561	TTACCTTACTGGACCGGGCAGGTTGGTGGAAAGGAATCAAGCAGTTCTTCCTCGCAAAA	1620
Db	567	LeuProTyrtirpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLys	586
Qy	1621	GCAGTGGACCATGAGGAAACCCAGTTCGCTATAGCAGCAGCGAAGTAAATCACCTGAGT	1680
Db	587	AlaValAspHisGluGlnThrProValArgTyraSerSerSerGluValAsnHisLeuSer	606
Qy	1681	CCAAGAGAAGTCCACACAGTGTGTCAGGCTGACTCTGCAGAGATGTCTCAGCCACTGGTA	1740
Db	607	ProArgIuValThrThrValLeuGlnAlaAspSerAlaGluTyraAlaGlnProLeuVal	626
Qy	1741	GGAGGAATTGTTGGTACATTCATCAAAAGATCTACCTTTAAACCAAGAGAGAAAGAA	1800
Db	627	GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGlu	646
Qy	1801	GCAGGCTATGCAGACCTTAGATCTCTACAACCTCACAGGGCAGGAAGTTATCATGCGCTAT	1860
Db	647	AlaGlyTyraAlaAspLeuAspProTyraAsnSerProGlyGlnIuValTyraHisAlaTyra	666
Qy	1861	GCTGAACCACTCCCAATTACGGGGCTGAGTATGCAACCCCAATCATCATGAGCATGTCA	1920

Db 487 ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 506
 QY 1381 CAACAACTCCAGTCTGATATCAGAAATACCTACCTGTAATGTAACCAAGAT 1440
 Db 507 GlnThrAlaSerProAlaSerProAlaSerProAlaSerProAlaSerProAla 526
 QY 1441 GTAGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 Db 527 ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuLeuLeu 546
 QY 1501 ATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 Db 547 IleLeuValCysAlaThrThrAlaGlnAlaSerSerAlaGluThrThrLeuThr 566
 QY 1561 TTACCTTACTGGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Db 567 LeuProThrTrpAspArgAlaGlyThrTrpLysGlyMetLysGlnPheLeuProAlaLys 586
 QY 1621 GCAGTGGACCATGAGGAACCCAGTTCGCTATAGCAGCAGGCAAGTAAATCACC 1680
 Db 587 AlaValAspHisGluThrProValArgTrpSerSerSerGluValAlaSerHisLeuSer 606
 QY 1681 CCAGGAGAACTCCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 Db 607 ProArgGluValThrThrValLeuGlnAlaAspSerAlaGluThrAlaGlnProLeuVal 626
 QY 1741 GGAGGAATGCTGGTACATCTCAAGATCTACCTTTAAACCAGGAAGAAAGAA 1800
 Db 627 GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluLysGlu 646
 QY 1801 GCAGGCTATGCAGCTAGATCTTCAACTCACAGGCGGCAAGTAAATCATGCTAT 1860
 Db 647 AlaGlyThrAlaAspLeuAspProTrpAsnSerProGlyGlnGluValThrHisAlaTrp 666
 QY 1861 GCTGAACCACTCCCAATTCAGGGGCTGAGTATGCAACCCCAATCATCATGACATGTC 1920
 Db 667 AlaGluProLeuProIleThrGlyProGluThrAlaThrProIleIleMetAspMetSer 686
 QY 1921 GGGACCCCAACATCTCAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
 Db 687 GlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsn 706
 QY 1981 CAACCTCCCCCTAGTGGGAACCTTCAATACACTCTCTCCAGGACTGACAGCTGCTCC 2040
 Db 707 GlnProProLeuValGlyThrThrAsnThrLeuLeuSerArgThrAspSerCysSer 726
 QY 2041 TCAGCCAGGCGCCAGTATGATACCCGGAAGCTGGGAAGCCAGCTCTACCTGCCCCAGAC 2100
 Db 727 SerAlaGlnAlaGlnThrAspThrProLysAlaGlyLysProGlyLeuProAlaProAsp 746
 QY 2101 GAATTGGTGACAGGTGGCCAGAGACACACAAGAGTATCAGGAGCAGGAAGGATGGG 2160
 Db 747 GluLeuValThrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGly 766
 QY 2161 GAATGTCATGCTTTTAAAGAAATCCCT 2187
 Db 767 GluCysAspValPheLysGluIleLeu 775
 RESULT 3
 Q912V2 PRELIMINARY; PRT; 769 AA.
 ID Q912V2
 AC Q912V2;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
 DE ESDN.
 GN Rattus norvegicus (Rat).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY;
 RX PubMed=11447234;
 RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
 RA Matsumori A., Sasayama S., Honjo T., Tashiro K.,
 RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
 RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
 RT Up-regulated after Vascular Injury.";
 RL J. Biol. Chem. 276:34105-34114(2001).
 CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 DR EMBL; AF387549; AAL30180.1; -.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR004043; LCCL_dom.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF00734; F5_F8_type_C; 1.
 DR PROSITE; PS01180; CUB; 1.
 KW DNA-binding; Transcription regulation.
 SQ SEQUENCE 769 AA; 83915 MW; 5B7936C8FA063A93 CRC64;
 Alignment Scores:
 Pred. No.: 5,91e-281 Length: 769
 Score: 3415.50 Matches: 634
 Percent Similarity: 93.22% Conservative: 40
 Best Local Similarity: 87.68% Mismatches: 46
 Query Match: 86.05% Indels: 3
 DB: 11 Gaps: 2
 US-10-060-830-2 (1-2190) x Q912V2 (1-769)
 QY 19 CTCCTACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 78
 Db 50 LeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 69
 QY 79 GGACACACGTACTAGCCCTGAGAGTGGAACTTACATCCATAACTACCCACAGACC 138
 Db 70 GlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerIleAsnThrProHisThr 89
 QY 139 TATCCCAACAGCAGCTGTTGTGAATGGAGATCCGTTAAAGATGGAGAGAGATTTCG 198
 Db 90 TyrProAsnSerThrValCysLysTrpGluLeuArgValLysThrGlyGluArgIleArg 109
 QY 199 ATCAATTTGGTGACCTTGACATTTGAGATTCGATCTGTCACITTAATTACTTGA 258
 Db 110 IleLysPheGlyAspPheAspIleGluAspSerAspTyrCysHisLeuAsnTyrLeuLys 129
 QY 259 ATTATTAATGGAATGAGTTCAGCAGAGAACTGAAATAGGCAATCTGCTGGGCTTG 318
 Db 130 IlePheAsnGlyIleGlyValSerArgThrGluIleGlyLysTyrCysGlyLeuGlyLeu 149
 QY 319 CAAATGAACCATTTCAATTTGAATCAAAAGCAATGAATCAATCATCTGTTTCATGAGTGA 378
 Db 150 GlnMetAsnGlnSerIleGluSerLysGlySerGluIleThrValLeuPheMetSerGly 169
 QY 379 ATCCATGTTTCTGGACGGGATTTTGGCCTCATCTCTCTTATAGATAAACAAGATCTA 438
 Db 170 IleHisAlaSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeu 189
 QY 439 ATTACTTGTTCGACACTGCATCCAAATTTTGGAACTCTAGTTCAGTACGATGCTGCCA 498
 Db 190 IleThrCysLeuAspThrValSerAsnPheLeuGluProGluPheSerLysTyrCysPro 209
 QY 499 GCTGGTGTCTGCTTCTTCTTCTGCTGAGATATCTGGAAACAATTCCTCATGGATATAGAT 558
 Db 210 AlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAsp 229
 QY 559 TCCTGCCCATTTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 618
 Db 230 SerSerProLeuCysMetAlaGlyIleHisAlaGlyValValSerAspValLeuGlyGly 249
 QY 619 CAAATCAGTGTCTAATTAGTAAGGTATTCCTATTTATGAAAGTCTTTGGCTAACAC 678

Db 250 GlnIleSerValValIleSerLysGlyThrProTyrTyrGluSerSerLeuAlaAsnAsn 269
 QY 679 GTCACATCTGGTGGGACACTTACTACAGTCTTTTACATTTTAAAGCAAGTGGATCT 738
 Db 270 ValThrSerMetValGlyTyrLeuSerThrSerLeuPheThrPheLysThrSerGlyCys 289
 QY 739 TATGGAACACTGGGATGGAGTCTGGTGTGATCGCGGATCTCTCAATAACACGATCATCT 798
 Db 290 TyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSer 309
 QY 799 GTGCTGGAGTGGACACACAGGCAAGAGACAGTGGAAACCCCAAAAGCCAGG 858
 Db 310 ValLeuGluThrAspHisMetGlyGlnGluAsnSerTyrLysProGluLysAlaArg 329
 QY 859 CTGAAACACCTGGACCCCTGGCTGCTTTGGCCACTGATGAATACAGTGGTTACAA 918
 Db 330 LeuArgLysProGlyProThrAlaAlaPheAlaThrAspGluHisGlnThrLeuGln 349
 QY 919 ATAGATTTGAATAAGCAAGAAATAACAGGCATTATACCACTGGATCCACCATGGTG 978
 Db 350 IleAspLeuAsnLysGluLysIleThrGlyIleValThrThrGlySerThrLeuIle 369
 QY 979 GAGCAATATTACTGTCTGCCCTACAGATCTCTGTACAGTATGATGGCAGAAATGG 1038
 Db 370 GluHisAsnTyrTyrValSerAlaTyrArgValLeuTyrSerAspGlyGlnLysTyr 389
 QY 1039 ACTGTGTACAGAGCCCTGGTGTGGCAGCAAGATATATTTCAAGGAACAAAGATTAT 1098
 Db 390 ThrValTyrArgGluProGlyAlaAlaGlnAspLysIlePheGlnGlyAsnLysAspTyr 409
 QY 1099 CACCAGGATGCTGTAATAACTTTTGGCCACCAATATTGCACGTTTATTAGAGTGAAT 1158
 Db 410 HisLysAspValArgAsnAsnPheLeuProPheIleAlaArgPheIleArgValAsn 429
 QY 1159 CCTACCAATGCGACAGAAATTCGATGAATAAGAGCTGCTCGATGTTCAGTTTATT 1218
 Db 430 ProValGlnThrGlnGlnLysIleAlaMetLysValGluLeuLysCysGlnPheThr 449
 QY 1219 CCTAAGGTGCTCCCAAACTTACTCACTCCCTCCCTCGGACAGCAATGACCTC 1278
 Db 450 LeuLysGlyArgLeuProLysLeuThrGlnProProProArgAsnSerAsnAsnLeu 469
 QY 1279 AAAACACTACAGCCCTCCAAATAGCCAAAGCTGCGCCCAAAATTTAGCAACCA 1338
 Db 470 LysAsnThrValHisProLysLeu-----GlyArgAlaProLysPheThrGlnAla 487
 QY 1339 CTACAACCTCGCAGTAGCAATGAATTTCTGCACAGACAGAACAACTGCGCATCTCT 1398
 Db 488 LeuGlnProArgSerArgAsnAspLeuProLeuLeuProAlaGlnThrThrAlaThrPro 507
 QY 1399 GATATCAGAAATACCTACCGTAACCTCAATGTAAACCAAGATGCTAGCGTGCAGTT 1458
 Db 508 AspValLysAsnThrThrValThrProSerValThrLysAspValAlaLeuAlaVal 527
 QY 1459 CTGTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1518
 Db 528 LeuValProValLeuValMetAlaLeuThrThrLeuIleLeuLeuValCysAlaTyr 547
 QY 1519 CACTGAGAAACAGAAAGAAACAACTGAAGCCACCTATGACTTACTTACTGGACCGG 1578
 Db 548 HisTyrArgAsnArgLysLysAlaGluGlyThrTyrAspLeuProHisTyrAspArg 567
 QY 1579 GCAGGTGTGGAAAGAAATGAAGCAGTTCCTCTGCAAAACAGTGGACCATGAGGAA 1638
 Db 568 AlaGlyTyrTyrLysGlyValLysGlnLeuLeuProAlaLysSerValGluHisGluGlu 587
 QY 1639 ACCCGAGTTCGCTATAGCAGCAGCGAAGTTAATCACTGAGTCCAGAGAGAGTCAACACA 1698
 Db 588 ThrProValArgTyrSerAsnSerGluValSerHisLeuSerProArgGluValThrThr 607
 QY 1699 GTCTGCAGGCTGACTCTGCAGAGTATGCTCAGCCACTGTGAGGAGGAATTTGTTGAC 1758
 Db 608 ValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThr 627

QY 1759 CTTTCATCAAGATCTACCTTTAAACCAAGAAAGAAAGCAGGCTATGACAGCCTA 1818
 Db 628 LeuHisGlnArgSerThrPheLysProGluGluGlyLysGluAlaSerTyrAlaAspLeu 647
 QY 1819 GATCTTTCAACTACCAAGGAGGAGGAGTTTATCATGCTATGCTGACCACTCCCAATT 1878
 Db 648 AspProTyrAsnAlaProValGlnGluValTyrHisAlaTyrAlaGluProLeuProVal 667
 QY 1879 ACGGGGCTGAGTATGCAACCCCAATCATCATGACATCTCAGGCACCCCACTTCA 1938
 Db 668 ThrGlyProGluTyrAlaThrProIleValMetAspMetSerGlyHisSerThrAlaSer 687
 QY 1939 GTTGTCTGACGCTCCACATCCACTTTTCAAGGCTCAGGGAACCAACCTCCCACTAGTG 1998
 Db 688 ValGlyLeuProSerThrSerThrPheArgThrAlaGlyAsnGlnProProAlaLeuVal 707
 QY 1999 GGAATTTCAATPACTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGCCAGTAT 2058
 Db 708 GlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerGlyGlnAlaGlnTyr 727
 QY 2059 GATACCCCAAGCTGGGAGGAGGCTTACCTGCCCCAGACGAATTTGGTGATCCAGGTG 2118
 Db 728 AspThrProLysGlyGlyLysProAla---AlaAlaProGluGluLeuValTyrGlnVal 746
 QY 2119 CCACAGAGCACACAAGATATCATCAGGAGGAGGAGGATGGGAATGTGATGTTTTTAA 2178
 Db 747 ProGlnSerThrGlnGluAlaSerGlyAlaGlyArgAspGluLysPheAlaPheLys 766
 QY 2179 GAAATCCTT 2187
 Db 767 GluThrLeu 769
 RESULT 4
 Q91ZV3 PRELIMINARY; PRT; 769 AA.
 AC Q91ZV3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
 GN ESDN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR;
 RX PubMed=11447234;
 RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
 RA Matsumori A., Sasayama S., Honjo T., Tashiro K.; Cloned from Vascular
 RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
 RT Up-regulated after Vascular Injury.
 RL J. Biol. Chem. 276:34105-34114(2001).
 CC 1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 DR EMBL; AF387548; AAL30179.1;
 DR MGD; MGI:1920629; ESDN.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR004043; LCCL_dom.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF00754; F5_F8_type_C; 1.
 DR PROSITE; PS01180; CUB; 1.
 KW DNA-binding; Transcription regulation.
 SQ SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;

Alignment Scores:
 Pred. No.: 3,73e-278 Length: 769
 Score: 3382.50 Matches: 627

QY 2119 CCACAGACACACAGAAGTATCAGGACGAGGATGGGAATGTGATGCTTTTAA 2178
 DB 747 ProGlnSerThrGlnGluLeuSerGlyAlaGlyArgSpGluLysPheAspAlaPheLys 766
 QY 2179 GAAATCCTT 2187
 DB 767 GluIleLeu 769
 RESULT 5
 ID Q14089 PRELIMINARY; PRT; 364 AA.
 AC Q14089;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 40.0 kDa protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shibata T.;
 RT "unpublished";
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL; D29810; BAA18909.1; -;
 DR InterPro; IPR000859; CUB_dom.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR004043; LCCL_dom.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF00754; FS_F8_type_C; 1.
 DR SMART; SM00042; CUB; 1.
 DR PROSITE; PS01180; CUB; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;

Alignment Scores:

Pred. No.: 3,73e-129 Length: 364
 Score: 1625.00 Matches: 342
 Percent Similarity: 94.81% Conservative: 5
 Best Local Similarity: 93.44% Mismatches: 14
 Query Match: 40.94% Indels: 11
 DB: 4 Gaps: 2

US-10-060-830-2 (1-2190) x Q14089 (1-364)

QY 70 GATGGATGGACACTGCTAGTCCCTGAGAGTGAACCTTACATCCATAAATAC 129
 DB 1 AspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerIleAsnTyr 20
 QY 130 CCACAGACCTATCCCAACAGCACTGTTGTGAATGGGAGATCCGTGAAGATGGGAGAG 189
 DB 21 ProGlnThrTyrProAsnSerThrValCysGluTrpGluIleArgValLysMetGlyGlu 40
 QY 190 AGATTCGCATCAAAATTTGGTGACTTTGACATTTGACATTTGATCTCTGCTCACTTAAT 249
 DB 41 ArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCysHisPheAsn 60
 QY 250 TACTTCAGATTTATATGAATTTGAGTGCAGTGCAGGACGAGTCAATAGGCAATGCTGGT 309
 DB 61 TyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyTyrCysGly 80
 QY 310 CTGGGTTCGAATGAACATTCATTAATGAATCAAAAGGCAATGAATGCAATGCTGCTTC 369
 DB 81 LeuGlyLeuGlnIleAsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPhe 100
 QY 370 ATGATGGGAATCCATGCTTTCTGGACGCGGATTTTGGCCCTCATCTCTGTATAGATAAA 429
 DB 101 MetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLys 120

QY 430 CA-AGATCTAATTAATCTGTTTGGACACTGCATCCCAATT-TTTTGGAACTGAGTTCACTA 487
 DB 121 GlnArgSerAsnTyrLeuPheGlyHisCysIleGlnPhePheTyrAsnLeuSerVal 140
 QY 488 AGTACTGC-CCAGCTGGTGTCTGCTTCTTCTTGTGTGAGATATCTGGAACAATTCCTCAT 546
 DB 141 SerThrAlaProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIleProHis 160
 QY 547 GGATATAGAGATTCCTCGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606
 DB 161 GlyTyrArgAspSerSerProLeuGlyMetAlaGlyValHisAlaGlyValValSerAsn 180
 QY 607 ACCTTGGCGGCAAACTAGTCTGTAAATAGTAAGGTATTCCTTATATGAAAGTTCT 666
 DB 181 ThrLeuGlyGlnIleSerValIleSerLysGlyIleProTyrTyrGluSerSer 200
 QY 667 TTGGGTACAACAGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 723
 DB 201 LeuAlaAsnAsnValThrSerValGlyHisLeuIleLeuGlnValPhePheThrPhe 220
 QY 724 AAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 782
 DB 221 LysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyGlyAspArgGlySerSer 240
 QY 783 AATAACAGCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 842
 DB 241 AsnAsnSerIle-ThrValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTyrPly 260
 QY 843 ACCCAAAAAGC-CAGGCTGAAAAACCTGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901
 DB 260 spTolLysSerGlnAlaGluLysThrTrpAlaLeuGly-AlaPheAlaThrAspG 280
 QY 902 ATACCACTGGTTACAAATAGATTGAATAAGAAAAGAAATAACAGCACTATATAACCA 961
 DB 280 LutyGlnTrpLeuGlnIleAspLeuAsnLysLysIleThrGlyIleIleThr 300
 QY 962 CTGGATCCACCATGGTGGACCAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1021
 DB 300 hrGlySerThrMetVal-SerThrIleThrMetCysLeuProThrGluSerCysThrVal 319
 QY 1022 ATGATGGGCAAAATGGACTGTGTACAGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 DB 320 MetMetGlyArgAsnGlyLeuCysThrGluSerLeuValValGluGlnAspLysIlePhe 339
 QY 1081 CAAGGAAACAAAGAT--TATCACCAGGATGCTGCTGAATAAATCTTTTGGCCACCAATATT 1137
 DB 340 GlnGlyAsnLysArgIleIleThrArgMetValArgAsnAsnPheLeuProIleIle 359
 QY 1138 GCACGTTTATT 1149
 DB 360 AlaArgLeuLeu 363
 RESULT 6
 ID Q9D9K5 PRELIMINARY; PRT; 251 AA.
 AC Q9D9K5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 170005P21Rik protein.
 GN ESDN OR 170005P21R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,


```
QY 124 AACTCCACAGACCTATCCCAACAGCACTGTTTGTGAATGGAGATCGGTGAAGATG 183
Db 124 11111111 1111111111 1111111111 111 111
QY 48 AsnThrProGlyThrTyProAsnThrValCysGluLysIleIleThrValProLys 67
Db 48 11111111 1111111111 1111111111 111 111
QY 184 GGAGAGAGTTCGCATCAAAATTTGGTACTTTGACATTTGAAGATTCATCTTGTCTAC 243
Db 184 11111111 1111111111 1111111111 111 111
QY 68 GlyLysArgLeuIleLeuArgLeuGlyAspLeuAsnIleGlu---SerLysThrCysAla 86
Db 68 11111111 1111111111 1111111111 111 111
QY 244 TTTAATTACTTGAGATTTATGAATGGAGTGGAGTGCAGCACTCAAAATAGCAATAC 303
Db 244 11111111 1111111111 1111111111 111 111
QY 87 SerAspThrLeuLeuPheSerSer-----AlaThrAspGlnThrGlyProThr 102
Db 87 11111111 1111111111 1111111111 111 111
QY 304 TGTGGTCTGGGTGCAATCAACCAATTCAATGAATCAAAAGGCAATGAATACACATG 363
Db 304 11111111 1111111111 1111111111 111 111
QY 103 CysGly---SerTrpAlaValProLysGluLeuArgLeuAsnSerAsnGluValThrVal 121
Db 103 11111111 1111111111 1111111111 111 111
QY 364 CTGTTTCATGAGTGAATCATGTTTCTGGACGCGGATTTTGGCCATCACTCTGTATA 423
Db 364 11111111 1111111111 1111111111 111 111
QY 122 LeuPheLysSerGlySerHisIleSerGlyArgGlyPheLeuLeuThrTyAlaSerSer 141
Db 122 11111111 1111111111 1111111111 111 111
QY 424 GATAACACAGACTAATTACTTGTGTTGGCACTGCATCCAAATTTTGGACCTGAGTTC 493
Db 424 11111111 1111111111 1111111111 111 111
QY 142 AspHisProAspLeuIleThrCysLeuGluArgGlySerHisTyPheGluGluLysTy 161
Db 142 11111111 1111111111 1111111111 111 111
QY 484 AGTAAGTACTGCCAGCTGTTGCTGCTCCTTTCTGAGATATCTTGGAAACAATTCCT 543
Db 484 11111111 1111111111 1111111111 111 111
QY 162 SerLysPheCysProAlaGlyCysArgAspIleAlaGlyAspIleSerGlyAsnThrLys 181
Db 162 11111111 1111111111 1111111111 111 111
QY 544 CATGATATAGAGATTCCTCGCAATTTGTCATGGCTGGTGTGCATCAGCAGGAGTGTCA 603
Db 544 11111111 1111111111 1111111111 111 111
QY 182 AspGlyTyArgAspThrSerLeuLeuCysLysAlaAlaIleHisAlaGlyIleIleThr 201
Db 182 11111111 1111111111 1111111111 111 111
QY 604 AACACGTTGGGCGGCAAAATCACTGTTGTAAATTAAGGATTCCTCTATTATGAAGT 663
Db 604 11111111 1111111111 1111111111 111 111
QY 202 AspGluLeuGlyGlyHisIleAsnLeuLeuGlnSerLysGlyIleSerHisTyGluGly 221
Db 202 11111111 1111111111 1111111111 111 111
QY 664 TCTTTGGCTAACACGCTCATCTGTTGGGACACTTACTACAAAGTCTTTTACATTT 723
Db 664 11111111 1111111111 1111111111 111 111
QY 222 LeuLeuAlaAsnGlyValLeuSerArgHisGlySerLeuSerGluLysArgPheLeuPhe 241
Db 222 11111111 1111111111 1111111111 111 111
QY 724 AGACAAAGTGGATTTATGGAACACTGGGGATGGAGTCTGTTGTGATCGCGGATCTCAA 783
Db 724 11111111 1111111111 1111111111 111 111
QY 242 ThrThr----- 243
Db 242 11111111 1111111111 1111111111 111 111
QY 784 ATACAGCATCATCTGTGTGGAGTGGAGTGCACACACAGGCAAGAGACAGTTGAAA 843
Db 784 11111111 1111111111 1111111111 111 111
QY 243 ----- 243
Db 243 11111111 1111111111 1111111111 111 111
QY 844 CCAAAAAAGCCAGCTGAAAAAACCTGGACCGCTTGGCTGCTTTTGGCCACTGATGAA 903
Db 844 11111111 1111111111 1111111111 111 111
QY 243 ----- 243
Db 243 11111111 1111111111 1111111111 111 111
QY 904 TACCAGTGGTTACAAATAGATTGTAATAGAAAAAAGAAAAATAACAGGCATTATAACCACT 963
Db 904 11111111 1111111111 1111111111 111 111
QY 243 ----- 243
Db 243 11111111 1111111111 1111111111 111 111
QY 964 GGATCCACCATGTTGGAGACAAATTAATATGTGTCTCCCTACAGAATCCTGTACAGTAT 1023
Db 964 11111111 1111111111 1111111111 111 111
QY 243 ----- 243
Db 243 11111111 1111111111 1111111111 111 111
QY 1024 GATGGCAGAAATGGACTGTGTACAGAGACGCTGGTGTGGAGCAAGATAAGATATTTCAA 1083
Db 1024 11111111 1111111111 1111111111 111 111
QY 244 ----- 246
Db 244 11111111 1111111111 1111111111 111 111
QY 1084 GGAACAAAGATTATCACCAGGATGTGCGTAATAAATTTTGGCCACCAATTATTGCACGT 1143
Db 1084 11111111 1111111111 1111111111 111 111
QY 246 ----- 246
Db 246 11111111 1111111111 1111111111 111 111
QY 1144 TTTATTAGATGAATCCTACCAATGCGACAGAAAAATGCCATGAAATGGAGCTGCTC 1203
Db 1144 11111111 1111111111 1111111111 111 111
QY 246 ----- 246
Db 246 11111111 1111111111 1111111111 111 111
```

RESULT 8

QY 691 GTGGGACACATATCTACAGTCTTTTACATTTAAGACAAGT----- 732
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 733 -----GGATGTTAT 741
Db 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProGluAsnPheGlnCysAsn 278
QY 742 GGAACACTGGGGATGGAGTCTGGTGATCGCGATCTCTCAATAACACAGCATCTGTG 801
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGlnIleSerAlaSer--- 297
QY 802 CTGGAGTGGACTGACACAGGACAGACAGTGGAAACCCCAAAAGCCAGGCTG 861
Db 298 -----ThrPheSerAspGlyArgTyrThrProGlnGlnSerArgLeu 311
QY 862 AAAAAACCTGGACCGCTTGGCTGCTTTGCCACTGATCAATACAGTGGTTACAATA 921
Db 312 HisGlyAspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnVal 331
QY 922 GATTGTAATAGGAAAAGAAATAACAGGATTATACCACTGGA-----TCCACCATG 975
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 976 GTGGAGCACATTAATGATGCTGCTGCCTACAGATCTGTACAGTATGATGGCAGAA 1035
Db 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuValSerThrAsnGlyGluAsp 371
QY 1036 TGGACTGTACAGAGCGCTGGTGTGGCAAGATAGATATTTCAAGGAAACAAAGAT 1095
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsp 389
QY 1096 TATCACCAGGATGTCGTATTAATATTTTCCACCAATATTACACGTTTTATTAGAGT 1155
Db 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1156 AATCCTACCCNATGGCAGCAGAAATTTGCCATGAAATGAGTCTCGGATGTCAGTTT 1215
Db 410 ArgProGlnThrThrPheLysLeuGlyIleAlaLeuArgLeuLeuPheGlyCys----- 427
QY 1216 ATTCTAAAGTCTGCTCTCAAACTTACTCAACTCTCTCGGAACAGCAATGAC 1275
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1276 CTCAAAACACTACAGCCCTCCAAAATAGCCAAAGTGTGCCCCCAAAATTT----- 1329
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTyr 460
QY 1330 -----ACGCAACCACTACACCTCGCAGTAGCAATGAATTTCTGCACACAGAA 1380
Db 461 SerProSerAlaAlaArgLeuValSerArgSerGlyTyrPheProArgAsnProGln 480
QY 1381 CAACAACACTGCCAGTCTGATATCAGAAATACCTCACTCAAAATGTAAACCAAGAT 1440
Db 481 AlaGlnProGlyGluGluTyrLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1441 GTA-----GCGTGCCTGCAGTCTGTGTCTGCTGCTG 1473
Db 501 ValIleIleGlnGlyAlaArgGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1474 GTCATGTCCTCACTACTCTCATCTCTATATTAGTGTGCTGCTGCACTGG----- 1524
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1525 -----AGAAACACAAAGAAAACCTGAAGGCACC-----TATGACTTACCTTACTGG 1572
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1573 GACCGGCGAGTGTGGTGAAGGAATGAAGCAGTTTCTCTCTGCAAAAGCAGTGGACCAT 1632
Db 559 -----AspIleArgArgPheAspPro----- 565

QY 1633 GAGGAACCCCACTTCGCTATATACACAGCAGGAAGTTAATCACCTGAGTCCACAGAGAAGTC 1592
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1693 ACCACAGTGTCTGAGGCTGACTCTGAGAGTAT-----GCTCAGCCACTGGTAGGAGGA 1746
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1747 ATTGTTGGTACACTTCATCAAGATCTACCTTTAAACCA-----GAAGAAGGA 1794
Db 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrPrometAspGluAspAla 623
QY 1795 AAGAAGCAGGCTAT-----GCAGACCTAGATCTCTCAACTCA 1833
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerGln 643
QY 1834 CCAGGCGAGGAAGTTATCATGCT-----ATGCTGAACCACTCCCA 1875
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTyrValTyrAspHisAlaLys 663
QY 1876 A-----TTACGGGGCTGAGTATGATGCAACCCCAATCATCATGAGCATGTGAGGCACCCCA-- 1930
Db 663 sTrpLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 1931 -----CAACTTCAG----- 1939
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArgGln 696
QY 1940 -----TTGTGTCAGCCCTCCACATCTCACTTTCAAGGCTACGGGGAACCAA 1983
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 1984 CTCCCCCACTAGTGGGAATTACATATCACTTCTCTC----- 2021
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2022 -----CAGACTGACAGCTGCTC-----CTCAGCCCGCCAG 2055
Db 731 lnValValArgGluAlaSerGlnGlnSerLysLeuLeuTyrValIle-ArgGluAspGln 750
QY 2056 TATGATACCCCGAAGCTGGGAAGCCAGGCTCTACCTGCCACAGCAATGGTGTACAG 2115
Db 751 GlySerGluTrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2116 GTCCACAGACACACAAGATATCAGACGAGGAAGGATGGGGAATGTGATGTT 2172
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGluIleSerIle 786
RESULT 10
Q8UVR0 PRELIMINARY; PRT; 919 AA.
AC Q8UVR0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Neuropilin-2al receptor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Herzog Y., Kalchelm C., Kahane N., Reshef R., Neufeld G.;
RT "Differential expression of neuropilin-1 and neuropilin-2 in arteries
and veins".
RL Mech. Dev. 0:0-0(2002).
DR EMBL; AF417235; AAL37723.1; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58.C.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.


```
Db 251 ThrAspLeuAlaValAlaLysAspGlyPheSerAlaGlnTyrTyrLeuIleGlnGlu 270
QY 733 -----GGATGTTATGAAACACTGGGATGAGTCTGGTGTGTCGCGGAT 777
Db 271 ValProGluAsnPhcGlnCysAsnValProLeuGlyMetGluSerGlyArgIleSerAsn 290
QY 778 CTTCAAAATACAGCATCTCTGCTGGTGGTGGTACTGACCACACAGGGGCAAGAACACT 837
Db 291 MetGlnIleSerAlaSerSer-----ThrTyrSerAspGlyArg 303
QY 838 TGGAAACCCAAAGACAGGCTGAAACAACTGGACCGCTTGGCTGCTTTGGCACT 897
Db 304 TrpThrProGlnGlnSerArgLeuAsnSerAspAsnGlyTrpThrProAsnValAsp 323
QY 898 GATGAATACCATGTTTCAATAGATTGTAATAGGAAAGAAATAACAGGCATTATA 957
Db 324 SerAsnLysGluTyrLeuGlnValAspLeuHisPheLeuThrValLeuThrAlaIleAla 343
QY 958 ACCACTGGA-----TCCACCATGTTGGAGCACATTTACTATGTCTGCTTACAGAATC 1011
Db 344 ThrGlnGlyAlaIleSerArgGluThrGlnAsnGlyTyrTyrValArgThrTyrLysLeu 363
QY 1012 CTGTACGTGATGATGGGAGAAATGGACTGTGTACAGAGAGCGCTGGTGGAGCAAGAT 1071
Db 364 GluValSerThrAsnGlyGluAspTrpMetMetTyrArgHisGly-----LysAsnHis 381
QY 1072 AGATATTTCAAGGAACAAAGATTATCACAGGATGTCGGTAATTAATCTTTTGGCCACCA 1131
Db 382 LysThrPheGlnAlaAsnGluAspAlaThrGluValValLeuAsnLysIleHisSerPro 401
QY 1132 ATTATGTCAGTGTATATAGATGAATCCTACCCATGCGCAGCAGAAATTCGCATGAA 1191
Db 402 ValLeuThrArgPheValArgIleArgProGlnSerTrpHisAsnGlyIleAlaLeuArg 421
QY 1192 ATGGAGCTGCTCGGATGTCAGTTTATTCCTAAAGTCGTCTCCAAACACTTACTCAA--- 1248
Db 422 LeuGluLeuTyrGlyCysArgIleThrAspSerProCysSerAsnLeuLeuGlyMetLeu 441
QY 1249 -----CCTCCACCTCCTCGGACACAGATAGCTCAAAACACACTACAGCCCT 1296
Db 442 SerGlyLeuIleProAspSerGlnIleSerAlaSerSerIleArgGlyTyrAspTrpSer 461
QY 1297 CCAAAATAGCCAAA-----GTCCTGCGCCCAAAATTTACGCAACCACTACACCT 1347
Db 462 ProSerMetAlaArgLeuValSerSerArgSerGlyTrpPheProArgIleProGlnAla 481
QY 1348 CGCAGTAGCAATGAATTT-----CCTGCACAGACAGAA----- 1380
Db 482 GlnProGlyGluTrpLeuGlnValAspLeuGlyValProLysAsnValLysGlyVal 501
QY 1381 -----CAAAACACTGCCAGTCTGTATATCAGAAATPACTACCGTAATCCAAATGTAACC 1434
Db 502 IleIleGlnGlyAlaArgGlyGlyAspSerValThrThrGluSerArgSerPheVal 521
QY 1435 AAAGATGATAGCGTGGCTCGACTTCTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1494
Db 522 LysLysPheLysValAla----- 527
QY 1495 ATTCTCATATTAGTGTGCTTGGCACTGGAGAAACAGAAAG----- 1536
Db 528 -----TyrSerMetAsnGlyLysAspTrpGluPheIleGln 539
QY 1537 -----AAAAAACTGAAGCACC-----TATGACTTACCTTAC 1569
Db 540 AspProLysThrMetGlnAlaLysLeuPheGluGlyAsnIleHisTyrAspIlePro--- 558
QY 1570 TGGGACCGGCGAGTTGGTGGAAAGGAATGAAGACAGTTTCTTCTCCGAAAGCAGTGGAC 1629
Db 559 -----GluIleArgArgPheAspPro----- 565
QY 1630 CATGAGGAACCCCACTTGCCTATACAGCAGCGCAAGATTAACTACCTGAGTCCAGAGAA 1689
```

```
Db 566 -----ValProAlaGlnTyrValArgValHisProGluArgTrpSerProAlaGly 582
QY 1690 GTCACACAGTCTGCTGAGGCTGACTGTCAGAGTAT-----GCTCAGCCACTGTGATAGA 1743
Db 583 IleGlyMetArgLeuGluValLeuGlyCysAspTrpThrAspValLysProThrAlaGlu 602
QY 1744 GGAATTTGTTGGTACACTTTCATCAAGATCTACCTTTAAACCA-----GAAGAA 1791
Db 603 ThrLeuValProThrLeuLysSerGluGluThrThrThrProTyrProThrAspGluGlu 622
QY 1792 GGAAGAAGACAGGCTATGATGACACCTAGATCTCTTACAACTACCCAGGGCAGGAAGTT--- 1848
Db 623 AlaThrGluCysGly-----AspSerCysGlyGluGluGluAsp 635
QY 1849 TATCATGCTCATGCTGAAC----- 1867
Db 636 PheHisLeuProAla--AsnPheAsnCysAsnPheAspLeuProGluAspLeuCysGlyTr 655
QY 1868 ----CACTCCCAATACGGGGCTG-----AGTATGCAACCCCAATCATCATGACATG 1917
Db 655 pSerHisAspSerAlaThrGlyTyrThrTrpSerPheGlnProAlaSerThrTrpMetG 675
QY 1918 TCAGGCGACCCCACTTTCAGTGTGCTAGCCCTCCACATCCACTTTCAGGCTACGGGG 1977
Db 675 yAsnSerGluProSerProGluThrLeuProAspAlaLysSerTyrLeuArgLeuGlnSe 695
QY 1978 AACCAACCTCCCTACTAGTGGAACTTACAATACATCTCTCCAGGACTACAGCTGC 2037
Db 695 rSerGly----- 697
QY 2038 TCCTCAGCCAGCCCGCAGTATGATACCCGAAAGCTGGGAGCGAGTCTACTGCCCCA 2097
Db 698 -----ArgArgGluSerGlnArgAlaArgLeuIleSerProThrIleTyrLeuProAr 715
QY 2098 GACGATTTGGTGT 2110
Db 715 gSerAlaValCys 719
RESULT 12
Q8R327 PRELIMINARY; PRT; 432 AA.
ID Q8R327 AC Q8R327;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to RIKEN CDNA 4631413K11 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC026771; RAH26771.1;
SQ SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD4917 CRC64;
Alignment Scores:
Pred. No.: 3.12e-26 Length: 432
Score: 412.00 Matches: 145
Percent Similarity: 33.08% Conservative: 71
Best Local Similarity: 22.21% Mismatches: 141
Query Match: 10.38% Indels: 296
DB: 11 Gaps: 19
US-10-060-830-2 (1-2190) x Q8R327 (1-432)
QY 289 GAAATAGGCAATPACTGTGTGCTGCGGTTCGAAATGAACATTCATTAATGAATCAAAAGGC 348
Db 27 GluLeuGlyProTyrCysGly---SerTrpAlaValProLysGluLeuArgLeuAsnSer 45
QY 349 AATGAATCACATTTGCTTTCATGTCAGTGAATCCATGTCGAGCGCGGATTTTGGCC 408
```

Db 46 AsnGluValThrValLeuPheLysSerGlySerHisIleSerGlyArgGlyPheLeuLeu 65
QY 409 TCATCTCTCTTTATAGATACAAAGATCTAATACCTTGTGTGGACACTGCATCCATTTT 468
Db 66 ThrTyrAlaSerSerAspHisProAspLeuIleThrCysLeuGluArgGlySerHisTyr 85
QY 469 TTGGAACTGAGTTAGTACTGCCAGCTGGTGTCTGCTCTCTTTTGTGTGAGATA 528
Db 86 PheGluGluLysTyrSerLysPheCysProAlaGlyCysArgAspIleAlaGlyAspIle 105
QY 529 TCTGGAACAATCTCTCATGATATAGAGATTCCTCCCATTTGTGATGCTGTGTGCAT 588
Db 106 SerGlyAsnThrLysAspGlyTyrArgAspThrSerLeuLeuLysAlaAlaIleHis 125
QY 589 GCAGGAGTAGTGTCAACACAGCTTGGCGGCGCAATCAGTGTGTATTAAGATAT 648
Db 126 AlaGlyIleThrAspGluLeuGlyGlyHisIleAsnLeuLeuGlnSerLysGlyIle 145
QY 649 CCTATTATGAAGTTCTTTGGCTAACACAGTCCACATCTGTGTGGACACTTATCTACA 708
Db 146 SerHisTyrGluGlyLeuLeuAlaAsnGlyValLeuSerArgHisGlySerLeuSerGlu 165
QY 709 AGCTTTTACATTTAAGACAAGTGTGATGTTATGAACACTGGGGATGGAGTCTGTGTG 768
Db 166 LysArgPheLeuPheThr 172
QY 769 ATCGGGATCCTCAATAACAGCATCATCTGTGCTGGAGTGGACTGACCACACAGGCAA 828
Db 172 172
QY 829 GAGACAGTTGGAAACCCAAAGCCAGGCTGAAAAACCTGGACCGCTTGGGTGCT 888
Db 172 172
QY 889 TTTGCCACTGATGAATACCAGTGTGTTACAAATAGATTGAATAAGAAAGAAATAACA 948
Db 172 172
QY 949 GGCATTAATACCACTGGATCCACCTGTTGGAGCACAATTACTATGTCTGTGCTACAGA 1008
Db 172 172
QY 1009 ATCCTGTACAGTGTATGGCAGAAATGGACTGTGTACAGAGAGCTGTGTGGAGCAA 1068
Db 173 175
QY 1069 GATAAGATATTTCAAGAAACAAGATATACACAGGATGTGGTAATAACTTTTGGCA 1128
Db 175 175
QY 1129 CCAATATTCCAGTGTATTAGTGAATCTTACCAATGGCAGAGAAATGGCCATG 1188
Db 175 175
QY 1189 AAAATGGAGTGTCTGGATGTCTGATTTATTCCTAAAGTGTCTCTCCAAACTTACTCAA 1248
Db 175 175
QY 1249 CCTCCACCTCTCGGAACGAATGACCTCAAAACACTACAGCCCTCCAAAATAGCC 1308
Db 175 175
QY 1309 AAAGTGTGCTCCCAAAATTTAGCAGACCACTACACCTCGCAGTAGCAATGAATTCCT 1368
Db 175 175
QY 1369 GCACAGACAGAAACAACACTGCCAGTCCGTGATATCAGAAATACTACCGTAACCTCAAAT 1428
Db 176 176
QY 1429 GTACCAAGAGTAGGCGTGGCTGCATGTTCTTGTCCCTGTGCTGTGCTGCATGGTCCACAT 1488
Db 177 IleThr---ThrValAlaIleProSerValIle-----PheIleAlaLeuLeuLeuThr 193

QY 1489 ACTCTCATCTCATATTAGTGTGTCTTGGCACTGGAGAAACAGAAAGAAACTGAA 1548
Db 194 GlyMetGlyIlePheAlaIleCys-----ArgLysArgLysLysLys 207
QY 1549 GGCACCTATGACTTACCTTAC-----TGGACCGCGCAGGTGTGGTGAAGGA 1596
Db 208 GlyAsn-----ProTyrValSerAlaAspAlaGlnLysThrGlyCysTyrLysGln 224
QY 1597 ATGAAGACATTTCTTCTCTGCAAAAGCAGTGGACCATCAGAAACCCCA---GTTCCGTAT 1653
Db 225 IleLysTyr-----ProPheAlaArgHis 232
QY 1654 ACACGACAGCAA-----GTTAATCACTCAGTCCAGAGAACTACC-----ACA 1698
Db 233 GlnSerThrGluPheThrIleSerTyrAspAsnGluLysGluMetThrGlnLysLeuAsp 252
QY 1699 GTGCTGCAGGCTCAGTCTGCAGAGTATGTCTCAGCCACTGTGTAGGAGAAATTTGTGTACA 1758
Db 253 LeuIleThrSerAspMetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrVal 272
QY 1759 CTTCAATCAAGATCTACCTTTAAACCA-----GAAGAA-----1791
Db 273 AlaArgLysGlySerThrPheArgProMetAspThrAspThrGluGluValArgValAsn 292
QY 1792 GGAAGAAGACGAGCTATGCAGACCTAGATCTTACAACTCACCAGGCGAGAGTTTAT 1851
Db 293 ThrGluAlaSerGlyHis-----TyrAspCysProHisArgProGlyArg 307
QY 1852 CATGCTATGCTGAACACTCCCAATTACGGGCTGAGTATGCAACCCCAATCATC---1908
Db 308 HisGluTyrAlaLeuProLeuThrHisSerGluProGluTyrAlaThrProIleValGlu 327
QY 1909 -----ATGCATATGCTAGGCGACCCCACTTACGTT 1941
Db 328 ArgHisLeuLeuArgAlaHisThrPheSerThrGlnSerGlyTyrArgValProGly---346
QY 1942 GGTACCCCTCCACATCCACTTTCAGGCTACGGGGAACCACTCCCCCACTAGTGGGA 2001
Db 347 ProArgProThrHisLysHisSerHisSerGlyGlyPheProProAlaThrGlyAla 366
QY 2002 ACTTACATACACTTCTCTCCAGCTGACAGCTGTCTCTCAGCCAGGCCAGTATGAT 2061
Db 367 ThrGlnIleGluSerTyrGlnArgProAlaSerProLysProValGlyGlyGlyTyrAsp 386
QY 2062 ACCCGAAAGCTGGGAGCCAGCTCTACCTGCCCGACAGCAANTTGTGTACAGGTCCTCA 2121
Db 387 LysPro---AlaAlaSerSerPheLeuAspSerArgAsp-----Pro 399
QY 2122 CAGAGCACACAGAAGTATCAGGACGAGGAGGATGGG 2160
Db 400 AlaSerGlnSerGlnMetThrSerGlyGlyAspGly 412

RESULT 13

Q96190 PRELIMINARY; PRT; 609 AA.
ID Q96190 AC Q96190; 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to neuropilin 1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007737; AA007737.1; -
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.

QY	712	CTTTTACATTTAAAGACAAGTGGATGCTTTATGAACAACATCGGGATGGAGTCTGGTGTGATC	771
Db	273	-----Phelys-----CysMetGluAlaLeuGlyMetGluSerGlyGluIle	286
QY	772	CGGGATCCTCAAAATAACAGCATCATCTGTGCTGGAGTGGAGCTGACACACACAGGGCAAGAG	831
Db	287	HisSerAspGlnIleThrAlaSerSer-----GlnTyrSer	298
QY	832	AACAGTTGGAAACCCCAAAAGCCAGCGCTGAAAAAACCTGGACCGCTTGGCGTCTTTT	891
Db	299	ThrAsnTrpSerAlaGluArgSerArgLeuAsnTyrProGluAsnGlyTrpThrProgly	318
QY	892	GCACGTGATGAATACACAGCTGGTTACAAATAGATTTTGAATAAGAGAAAGAAAATAACAGGC	951
Db	319	GluAspSerTyrArgGluTrpIleGlnValAspLeuGlyLeuLeuArgPheValThrAla	338
QY	952	ATTATACCACTGGA-----TCCACCATGGTGGAGCACAAATACTATGTGCTGCCTAC	1005
Db	339	ValGlyThrGlnGlyAlaIleSerLysGluThrLysLysLysTyrValLysThrTyr	358
QY	1006	AGATCTCTGTACAGTGTATGGCGAGAAATGGACTGTGTACAGAGAGCGCTGGTGGAG	1065
Db	359	LysIleAspValSerSerAsnGlyGluAspTrpIleThrIleLysGluGly-----Asn	376
QY	1066	CAGATAAGATATTTCAGAGAACAAAGATATTATCACAGGATGTGCGTAATAACTTTTG	1125
Db	377	LysProValLeupheGlnGlyAsnThrAsnProThrAspValValAlaValPhePro	396
QY	1126	CCACCAATTATTCAGCGTTTATTACAGTGAATCTTACCCCAATGGCAGACAGAAATGCC	1185
Db	397	LysProLeuIleThrArgPheValArgIleLysProAlaThrTrpGluThrGlyIleSer	416
QY	1186	ATGAAATGGAGCTGCTCGAGTGTACGTTTATTCCTAAAGGTCGCTCCCAAAACTTACT	1245
Db	417	MetargPheGluValTyrGlyCys-----LysIleThr	427
QY	1246	CAACCTCCA 1254	
Db	428	AspTyrPro 430	
RESULT 14			
QY	961H5	PRELIMINARY;	PRT; 644 AA.
ID	AC	Q96IH5;	
AD	Q96IH5;		
DT	01-DEC-2001	(TrEMBLrel. 19, Created)	
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)	
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)	
OS	Neuropilin 1.		
DE	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	TISSUE=KIDNEY;		
RA	Strausberg R.;		
RL	EMBL: BC007533; A3H07533.1; -		
DR	Submitted (WAY-2001) to the EMBL/GenBank/DBJ databases.		
DR	InterPro: IPR000859; CUB_domain.		
DR	InterPro: IPR000421; FA58_C.		
DR	Pfam: PF00431; CUB; 2.		
DR	Pfam: PF00754; F5_F8_type_C; 2.		
DR	PROSITE; PS01180; CUB; 2.		
DR	PROSITE; PS01285; FA58C1; UNKNOWN_2.		
DR	PROSITE; PS01286; FA58C2; UNKNOWN_2.		
SQ	SEQUENCE 644 AA; 71907 MW; 9204B6F5C3CD2A21 CRC64;		
Alignment Scores:			
Pred. No.:	3 3e-25	Length:	644
Score:	400.50	Matches:	130
Percent Similarity:	42.33%	Conservative:	66
Best Local Similarity:	28.08%	Mismatches:	178
Query Match:	10.09%	Indels:	89

Alignment Scores:	
Pred. NO.:	3.3e-25
Score:	400.50
Length:	644
Percent Similarity:	42.33%
Matches:	130
Best Local Similarity:	28.08%
Conservative:	66
Mismatches:	178
Query Match:	10.09%
Indels:	89


```
; REFERENCE /DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 901 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-936-135-22

Alignment Scores:
Pred. No.: 1,07e-30 Length: 901
Score: 414.50 Matches: 205
Percent Similarity: 36.59% Conservative: 110
Best Local Similarity: 23.81% Mismatches: 336
Query Match: 10.44% Indels: 210
DB: Gaps: 37

US-10-060-830-2 (1-2190) x US-08-936-135-22 (1-901)
QY 1 ATGCCTCTGTTCCCTCGTCCTTACTTGTCCTGCTCCTGCTC-----GAGGACGCT 54
Db 1 MetAspMetPheProLeuThrTrpValPheLeuAlaLeuTyrrPheSerGlyHisGluVal 20
QY 55 GGAGCCCAAGAAGGTGATGGATGTGCACACACTGTACTAGGCCCTCGAGAGTGGAAACCCTT 114
Db 21 ArgSerGlnGlnAspProProCysGlyGlyArgProAsnSerLysAspAlaGlyTyrlle 40
QY 115 ACATCCATAAAGTACCACAGACCTATCCCACAGCACACTGTTTGTAATGGAGATCCGCT 174
Db 41 ThrSerProGlyTyrrProGlnAspTyrrProSerHisGlnAsnCysGluTrpIleValTyrr 60
QY 175 GTA---RAGATGGGAGAGAGATTGCATCAAAATTTGGT---GACTTTTGACATTGAAGAT 228
Db 61 AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGluLys 80
QY 229 TCTGATTCTTTGCTACTTTAAATTTACTTTGAGAAATTTATAATGGAATTTGGAGTCAGCAGA 288
Db 81 Hisasp--CysLysTyrrAspPheIleGluIleArgAspGlyAspSerGluSerAlaAsp 99
QY 289 GAATAGGCAAATFACTGTGTCTGGGTGTCAAATCAAC-----CATTCATTTGAA 339
Db 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProThrIleIle 114
QY 340 TCAAAAGSCAATGAATCAATGCTGTTTCATCAGTGAATCCATGTTTCTCGGACGCGGA 399
Db 115 SerSerGlySerValLeuTyrrIleLysPheThrSerAspTyrrAlaArgGlnGlyAlaGly 134
QY 400 TTTTGTGGCCTCATACTGCTTAGATATAACAAGATCTAATTACTTGTGTTGGACACTGCA 459
Db 135 PheSerLeuArgTyrrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 450 TCCAATTTTTGGAACTGAGTTCAGTAAGTACTGCCACAGCTGTTGTCGTGCTCTCTTTT 519
Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
QY 520 GCTGAGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTCGCCATTGTGTCATGGCT 579
Db 166 ProGlu-----LysTyrrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 580 GGTGTGCATGCAGGAGTAGTGTCAACACAGCTGTGGCGGCCAAATCAGT----- 627
Db 182 LysProArgMetGluIleIleLeuGlnPheLeuThrPheAspLeuGluHisaspProLeu 201
QY 628 -----GTTGTATATTAGTAAGGTATTCC----- 651
Db 202 GlnValGlyGluGlyAspCysLysTyrrAspTrpLeuAspIleTrpAspGlyIleProHis 221
QY 652 -----TATTATGAAGTCTTTTGGCTAACACAGTCACATCTGTG 690
```

RESULT 2
US-08-936-135-22
; Sequence 22, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627

Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSerSer 241
QY 691 GTGGGACACTTATCTACAGTCTTTTACATTTAGACAGT----- 732
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 733 -----GGATGTTTAT 741
Db 259 GlyPheSerAlaArgTyrThrLeuIleHisGlnGluProProGluAsnPheGlnCysAsn 278
QY 742 GGAACACTGGGATGGAGTCTGGTGTGATCGCGGATCTCAATAAACACATCATCTGTG 801
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer--- 297
QY 802 CTGGAGTGGACTGACACACAGCGGCAAGAGACAGTGTGAACCCCAAAAACCCAGGCTG 861
Db 298 -----ThrPheSerAspGlyArgTyrThrProGlnGlnSerArgLeu 311
QY 862 AAAAAACCTGGCGCTTGGCTGCTTTTGGCACTGATGAATACCAAGTGGTTACAAATA 921
Db 312 HisGlyAspAsnGlyTyrThrProAsnLeuAspSerAsnLysGlyIleVal 331
QY 922 GATTGTAATAAGAAAGAAATAACAGCATTAATAACCACTGGA-----TCCACCATG 975
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 976 GTGGACACAANTTACTATGTCTGCTGCTACAGATCTCTACAGTATGATGGCAGAAA 1035
Db 352 ThrGlnLysGlyTyrThrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1036 TGGACTGTCTACAGAGAGCTGTGTGGAGCAGAGATTAAGATATTTCAAGGAACAAGAT 1095
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsp 389
QY 1096 TATCACAGGATGTGGTGAATAACTTTTGGCCACAATATTATTCAGCTTTTATAGATG 1155
Db 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1156 AATCCTACCAATGGCAGAGAAAATTTGCCATGAAATGGAGTGTCTCGGATGTCAGTTT 1215
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1216 ATTCCTAAAGGTGCTCTCCAAACCTTACTCAACCTCCACCTCTCTCGGAACCAATGAC 1275
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1276 CTCAAAAACACTACAGCCCTCCAAAATAGCCAAAGTCTGCGCCCAAAATTT----- 1329
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerThrArgGluTyrLeuTyr 460
QY 1330 -----ACGCAACCACTACACCTCGCAGTAGCAATGAATTTCTTCACACAGAGAA 1380
Db 461 SerProSerAlaAlaArgLeuValSerArgSerGlyTyrPheProArgAsnProGln 480
QY 1381 CAAACAACCTGCCAGTCTGATATCAAAATACTACCGTAACCTCAAAATGAACCAAGAT 1440
Db 481 AlaGlnProGlyGluGlnThrPleuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1441 GTA-----GCGCTGGCTGCAGTCTTCTGCTCCCTGTGCTG 1473
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValAlaArgAlaPhe 520
QY 1474 GTCATGGTCTCACTACTCTCATTCATATAGTGTGCTTGGCAGCTGG----- 1524
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1525 -----AGAAACAAGAAAGAAATAAGAGGCACC-----TATGACTTACCTTACTG 1572
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1573 GACCGGCGAGTTGGTGGAAAGGAATGAAGCAGTTTCTTCCTCAAAAGCAGTGGACCAT 1632
Db 559 -----AspIleArgArgPheAspPro----- 565

QY 1633 GAGGAAACCCAGTTCGCTATAGCAGCAGCGAAGTTAATCACCTGAGTCCACAGAAGTC 1692
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTyrPheProAlaGlyIle 583
QY 1693 ACCACAGTCTGCAGCTGACTCTGCAGAGTAT-----GCTCAGCAGCTGTAGGAGGA 1746
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTyrThrAspSerLysProThrValGluThr 603
QY 1747 ATTGTTGGTACACTTCATCAAGATCTACTCTTAAACCA-----GAAGAGGA 1794
Db 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
QY 1795 AAGAAGCAGGCTAT-----GCACACCTAGATCTCTACAACTCA 1833
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerG1 643
QY 1834 CCAGGCGAGAAAGTTTATCATGCT-----ATGCTGAACCACTCCCA 1875
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTyrValTyrAspHisAlaL 663
QY 1876 A---TTACGGGCGCTGAGTATGCAACCCCAATCATCATGACATGTCAGGCGACCCCA-- 1930
Db 663 strpLeuArg-----SerThrTyrIleSerSerAlaAsnProAs 676
QY 1931 -----CAACTTCAG----- 1939
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArg1 696
QY 1940 -----TTGTGTCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACAA 1983
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 1984 COTCCCCCACTAGTGGGAATACAAATACACTTCTCTC----- 2021
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2022 -----CAGGACTGACAGCTGCTC-----CTCAGCCCGGCCCCAG 2055
Db 731 lnValValArgGluAlaSerGlnGluSerLysLeuLeuTyrValIle-ArgGluAspGln 750
QY 2056 TATGATACCCGAAAGCTGGGAAGCCAGGCTACCTGCCCGACAGCAATTTGGTGTACCAG 2115
Db 751 GlySerGluTyrLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2116 GTCCACAGACGACACAGAAGTATCAGGACAGGAGGATGGGATGGGATGTGATGTT 2172
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGluIleSerile 786

RESULT 3

US-08-936-135-24

; Sequence 24, Application US/08936135

; Patent No. 6054293

; GENERAL INFORMATION:

; APPLICANT: Tessier-Lavigne, Marc

; APPLICANT: He, Zhigang

; TITLE OF INVENTION: Semaphorin Receptors

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 75 DENISE DRIVE

; CITY: HILLSBOROUGH

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94010

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/936,135

FILING DATE: 435
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC97-288-2
TELEPHONE: (650) 343-4341
TELEPHONE: (650) 343-4342
TELEPHONE: (650) 343-4342
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 906 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-10-936-135-24

Alignment Scores:
Pred. No.: 1,07e-30 Length: 906
Score: 414.50 Matches: 205
Percent Similarity: 36.59% Conservative: 110
Best Local Similarity: 23.81% Mismatches: 336
Query Match: 10.44% Indels: 210
DB: 3 Gaps: 37

US-10-060-830-2 (1-2190) x US-08-936-135-24 (1-906)

QY	1	ATGCCTCTGTCCTCCCTGCTTACTTGTCTGCTGCTGCTGCTC-----GAGGACGCT	54
DB	1	MetAspMetPheProLeuThrTrpValPheLeuAlaLeuTyPheSerGlyHisGluVal	20
QY	55	GGAGCCAGCAAGGTGATGGATGGGACACACTGTACTAGCCCTGAGAGTGGACCCCT	114
DB	21	ArgSerGlnGlnAspProProCysGlyArgProAsnSerLysAspAlaGlyTyrIle	40
QY	115	ACATCCATAACTACCCAGACACTATCCCAACAGCACTGTTTGTGAATGGAGATCCGT	174
DB	41	ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnGlyTrpIleValTyr	60
QY	175	GTA---AGATGGGAGAGAGTTCGATCAAAATTTGCT---GACTTTGACATTCAAGAT	228
DB	61	AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGluLys	80
QY	229	TCTGATTCTTCTACTTTAATTAATGAGATTTATTAATGGAATTTGAGTACAGCAACT	288
DB	81	HisAsp---CysLysTyrAspPheIleGluLeuArgAspGlyAspSerGluSerAlaAsp	99
QY	289	GAATAGCAATAACTGTGGTCTGGGTTGCAATGAAC-----CATTCATTTGAA	339
DB	100	LeuLeuGlyHisCysGly-----AsnIleAlaProProThrIleIle	114
QY	340	TCAAAGGCAATGAATACACATTTGCTGTCATGAGTGGAAATCCATGTTCTCGACCGGA	399
DB	115	SerSerGlySerValLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly	134
QY	400	TTTTTGGCCTCATCTCTGTTATAGATAAACAAGATCTAATTAATCTGTTTGGACACTGA	459
DB	135	PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer	150
QY	460	TCCAATTTTTTGGAACTGAGTTCAGTAAGTACTCCAGCTGGTGTGCTGCTCTCTTTT	519
DB	151	LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe	165
QY	520	GCTGAGATATCTGGACAAATTCCTCATGGATATAGAGATTCCTGCCATTTGTCATGGCT	579
DB	166	ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla	181
QY	580	GGTGTGCATCAGGAGTAGTTCACACACAGCTGGCGGCCCAATACAGT-----	627
DB	182	LysProArgMetGluIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu	201
QY	628	-----GTTGTAATTAGTAAAGGATATCCC-----	651

Db	541	ProArgThrGlnGlnThrIysLeuPheGluGlyAsnMetHisTyrAspThrPro-----	558
QY	1573	GACCGGCAGGTTGGTGGAAAGGAATGAAGCAGGTTTCTCCTCGCAAAAGCAGTGGACCAT	1632
Db	559	-----AspIleArgArgPheaspPro-----	565
QY	1633	GAGGAACCCCGATTGCGCTATAGCAGCAGCGAAGTTTAACCTACGTGATCGCAAGAGATC	1692
Db	566	-----ValProAlaGlnTyrValArgValTyrProGluArGtrTpSerProAlaGlyIle	583
QY	1693	ACCACAGTGTCCAGGCTGACTCTGCACAGTAT-----GCTCAGCCACTGGTAGGAGGA	1746
Db	584	GlyMetArgLeuGluValLeuGlyCysAspIrpThrAspSerLysProThrValGluThr	603
QY	1747	ATTGTTGGTACACTTCATCAAAATCTACCTTTAAACA-----GAAGRAGGA	1794
Db	604	LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla	623
QY	1795	AAAGAGCAGGCTAT-----GCAGACTTAGATGCTTACACTCA	1833
Db	624	ThrGluCysGlyGluAsnCysSerPheGluAspAspLysAspLeuGln-LeuProSerG1	643
QY	1834	CCAGGCGAGGAAGTTTATCATGCC-----ATGCTGAACCACTCCCA	1875
Db	643	yPheAsnCysAsnPheaspPheProGluGluThrCysGlyTrpValTyrAspHisAlaLy	663
QY	1876	A-----TTACGGGCGCTGAGTATGCAACCCCAATCATCATGGACATGTGAGGCAACCCCA--	1930
Db	663	stIpLeuArg-----SerThrTrpIleSerSerAlaAsnProAs	676
QY	1931	-----CAACTTCAG-----	1939
Db	676	nAspA-gThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArgG1	696
QY	1940	-----TTGGTCAGGCGCTCCACATCCACTTTCAAGGCTACGGGGGAACCA	1983
Db	696	uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA	711
QY	1984	CTCTCCCGCACTAGTGGGAACCTTACAATACACTTCTCTC-----	2021
Db	711	rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyHisGlyValAlaLeuG	731
QY	2022	-----CAGGACTGCACCTGCTC-----CTAGCCCGAGGCCAG	2055
Db	731	lnValValArgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGln	750
QY	2056	TATGATACCCCGAAGACTGGGAAGCAGGTCTACTCTGCCCGACGAATGGTGTACCA	2115
Db	751	GlySerClutrpIyHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln	769
QY	2116	GTGCCACAGACACACAGAAGTATCAGGACGAGGAGGATGGGAATGTGATGTT	2172
Db	770	Ile-----ValPheGluGlyValIleGlyLysArgSerGluGluIleSerIle	786

RESULT 4

US-08-936-135-8

; Sequence 8, Application US/08936135

; Patent No. 6054293

GENERAL INFORMATION:

; APPLICANT: Tessier-Lavigne, Marc

APPLICANT: He, zh

APPLICANT: Chen,

; TITLE OF INVENTION: Semaphorin Receptors

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

STREET: 75 DENISE DRIVE

; CITY: HILLSBOROUGH

STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94010

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk

```

QY 580 GGTGTCATGACAGAGTACGTCAACACAGCTGTGGGGCGCAATCACTG----- 627
Db 182 LysProArgMetGluIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu 201
QY 628 -----CTTGTAATTAGTAAGCTATCCG--- 651
Db 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis 221
QY 652 -----TATTATGAAGTCTTGGCTAACACAGTCACATCTGTG 690
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSer 241
QY 691 GTGGGACACTTACTACAAGCTTTTACATTAAGACAAGT----- 732
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 733 -----GGATGTTAT 741
Db 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProProGluAsnPheGlnCysAsn 278
QY 742 GGAACACTGGGATGGAGTCTGTGTGATCGCGGATCTCTCAATAACAGCATCATCTGTG 801
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGlnIleSerAlaSer--- 297
QY 802 CTGGAGTGGACTGACACACAGCGGCAAGAGAACAGTGGAAACCCAAAGCCAGCGCTG 861
Db 298 -----ThrPheSerAspGlyArgTrpThrProGlnGlnSerArgLeu 311
QY 862 AAAAAACCTGGACCGCTTGGCTGCTTTGCCACTGATGATACCATCGTGTACAATA 921
Db 312 HisGlyAspAsnGlyTrpThrProAsnLeuAspSerAsnLysGlyTyrLeuGlnVal 331
QY 922 GATTGTAATAGGAAAAAATACAGGCATTATACCACTGGA-----TCCACCATG 975
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 976 GTGGAGCACAATTACTGTCTGCTGCCTACAGATCCTGTACAGTGTATGGGCGAGAA 1035
Db 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1036 TGGACTGTGTACAGAGCGCTGGTGGAGCAAGATAGATATTCAAGGAAACAAAGAT 1095
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsnAsp 389
QY 1096 TATCACCAGATGCGTAATACTTTTGGCCACCAATATTGTCAGCTTTATTAGATG 1155
Db 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1156 AATCCTACCCAAATGGCAGCAGAAATGCCATGAAATGGAGCTGCTCGGATGTCAGTTT 1215
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1216 ATTCTAAGCTCGCTCCCAAACTTACTCAACCTCCACCTCCTCGGAACCAATGAC 1275
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1276 CTCAAAAACACTACAGCCCTCCAAAAATAGCCAAAGGTGCTGCCCAAAATTT----- 1329
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTrp 460
QY 1330 -----ACGCAACCACTACAACTCGCAGTACGAATGAATTTCTTCGACAGACAGAA 1380
Db 461 SerProSerAlaAlaArgLeuValSerArgSerGlyTrpPheProArgAsnProGln 480
QY 1381 CAAACAACTGCCAGTCCGTGATATCAAGAAATACTACCGTAACCTCCAAATGTACCAAGAT 1440
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1441 GTA-----GGCTGGCTGCGAGTCTTCTCCCTGCTGCTG 1473
Db 501 ValIleIleGlnGlyAlaArgGlyGlyaspSerIleThrAlaValAlaGluAlaArgAlaPhe 520
QY 1474 GTCATGGTCTCCTCACTACTCTCATTTCTCATATTAGTGTGTGCTTGGCACTGG----- 1524

Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1525 -----AGAAACAGAAAGAAAAAACTGAAGGCACC-----TATGACTTACTTACTTGG 1572
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1573 GACCGGCAGGTGGTGGAAAGTAAGTAAAGCAGTTCTTCTCTGCAAAAGCAGTGGACCAT 1632
Db 559 -----AspIleArgArgPheAspPro----- 565
QY 1633 GAGGAACCCAGTTCGTATAGCAGCAGCGAGATTAATCACCTGAGTCCAAAGAGAAGTC 1692
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1693 ACCACAGTGTGTCAGGCTGACTCTGTCAGAGTAT-----GCTCAGCCACTGTGTAGGAGA 1746
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1747 ATTGTTGTACACTTCATCAAGATCTACCTTTAAACCA-----GAAGAAGGA 1794
Db 604 LeuGlyProThrValLysSerGluGluThrThrProTyrProMetAspGluAspAla 623
QY 1795 AAAGAAGCAGCTAT-----GCAGACCTAGATCTTCAACACTCA 1833
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerG1 643
QY 1834 CCAGGCGAGGAAGTTTATCATGCT-----ATGCTGAACCACTCCCA 1875
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTrpValTyrAspHisAlaLy 663
QY 1876 A---TTACGGGCGCTGAGTATGCAACCCCAATCATCATGACATGTGAGGCCACCCCA-- 1930
Db 663 strLeuArg-----SerThrTrpLysSerSerAlaAsnProAs 676
QY 1931 -----CAACTTCAG----- 1939
Db 676 nasArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArgG1 696
QY 1940 -----TTGGTCAGCCCTCCACATCCACTTTCAGGCTACGGGNAACCAA 1983
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 1984 CCTCCCCACTAGTGGGAACCTTACANTACTCTCTC----- 2021
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2022 -----CAGGACTGACAGCTGCTC-----CTCAGCCAGCCGAC 2055
Db 731 InValValArgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2056 TATGATACCCGAAAGCTGGGAGCCAGGTCTACTGCTCCCGCAGACGAATTTGTGTACCAG 2115
Db 751 GlySerGluTrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2116 GTGCCACAGACACACAGACTATCAGGACGAGGAGGATGGGAATGTGATGTT 2172
Db 770 ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGluIleSerIle 786
RESULT 5
US-08-936-135-10
; Sequence 10, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH

Db 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 460 TCAATTTTGGAACTGAGTTCAGTAACTAGTACTCCAGCTGGTGTCTGCTCTTTT 519
Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
QY 520 GCTGAGATATCTGGAACAATCTCTCGATATAGATATCTCGCCATTGTGCTGCT 579
Db 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 580 GGTGTGATCGCAGGATAGTGTCAACACGTTGGCGGCCAAATCAGT----- 627
Db 182 LysProArgMetGluIleLeuGlnPheLeuThrPheAspLeuGluHisProLeu 201
QY 628 -----GTTGTAATAGTAAAGGATTATCC----- 651
Db 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis 221
QY 652 -----TATTATGAAAGTCTTTGGCTTAACAACGTCACATCTGTG 690
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSer 241
QY 691 GTGGGACACTTATCAACAAGTCTTTTACATTTAAGACAAGT----- 732
Db 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 733 -----GGATGTTAT 741
Db 259 GlyPheSerAlaArgTyrTrpLeuIleHisGlnGluProProGluAsnPheGlnCysAsn 278
QY 742 GGAACACTGGGAGTGGTGTGTCGGGATCTCTCAATTAACAGCATCATCTGTG 801
Db 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSer 297
QY 802 GTGGAGTGGACTGACACACAGCGGCAAGAACAGTGTGGAACCCCAAAAGCCAGGCTG 861
Db 298 -----ThrPheSerAspGlyArgTyrProGlnGlnSerArgLeu 311
QY 862 AAAAAACCTGGACCGCTGGGCTGCTTTGGCACTGATGATACAGCTGGTTACAATA 921
Db 312 HisGlyAspAsnGlyTrpThrProAsnLeuAspSerAsnLysGluTrpLeuGlnVal 331
QY 922 GATTGTAATGAAGAAAGAAATAACAGGCATTATAACACCTGGA-----TCCACCATG 975
Db 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 976 GTGGACACAATATCTATCTGCTGCTACAGAATCCTGTACAGTATGATGGCAGAAA 1035
Db 352 ThrGlnLysGlyTyrTrpValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1036 TGGACTGTCTACAGACAGCTGCTGTGGAGCAAGATAGATATTTCAAGGAACAAGAT 1095
Db 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsp 389
QY 1096 TATCACAGGATGGTGGTAAATACTTTTGGCCACCAATATTGTCACGTTTATTAGAGTG 1155
Db 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
QY 1156 AATCCTTACCAATGGCAGAGAAAATTGCAATGGAATGGAGCTGCTCGGATGTCAGTTT 1215
Db 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1216 ATTCCTAAAGTGCTCCTCCAAACTTACTCAACCTCCACCTCCCTCGGGAACAGCAATGAC 1275
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1276 CTCAAAACACTACAGCCCTCCAAAATAGCCAAAGTCTGGTCCCAAAATTT----- 1329
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerThrArgGluTyrLeuTrp 460
QY 1330 -----ACGCAACCACTACAACCTCGCAGTAGCAATGAAATTTCTCTCCACAGACAGAA 1380
Db 461 SerProSerAlaAlaArgLeuValSerArgSerArgSerGlyTrpPheProArgAsnProGln 480

QY 1381 CAACAACCTGCCAGTCTGATATATCAGAAATACTACGTTAACTCCAAATGTACCAAGAT 1440
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1441 GTA-----GGCTGGCTGCAGTCTCTCTCCCTGTGCTG 1473
Db 501 ValIleIleGlnGlyAlaArgGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1474 GTCATGTGCTCTACACTCTCATTTATAGTGTGTGTGGCAGTGG----- 1524
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1525 -----AGAAACAGAAAGAAAAGTGAAGCACC-----TATGACTTACTTACTGG 1572
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro 558
QY 1573 GACCGGCGAGTGTGGTGAAGGAATGAAGCAGTCTTCTCTGCAAAAGCAGTGGACCAT 1632
Db 559 -----AspIleArgArgPheAspPro----- 565
QY 1633 GAGGAACCCAGTTCGCTATAGCAGCAGCAGGAAGTTAATCACTGAGTCCACAGAGATC 1692
Db 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1693 ACCAGTGTCTGACGCTGCTCTGCAGAGTAT-----GCTCAGCCACTGTTAGGAGGA 1746
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1747 ATTGTGTGTACTACTCATCAAGATCTACTTTAAACCA-----GAAGAAGGA 1794
Db 604 LeuGlyProThrValLysSerGluGluThrThrProTyrProMetAspGluAspAla 623
QY 1795 AAGAAGCAGGCTAT-----GCAGACTAGATCTTACAACTCA 1833
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspAspLysAspLeuGln-LeuProSerG 643
QY 1834 CCAGGCGAGGAGTGTATCATGCT-----ATGCTGAACCACTCCCA 1875
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTrpValTyrAspHisAla 663
QY 1876 A-----TTACGGGGCCTGAGTATGCAACCCCAATCATCATGAGCATGTCAGGCGCACCCA-- 1930
Db 663 strpLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 1931 -----CAACTTCAG----- 1939
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArg 696
QY 1940 -----TTGGTCAGCCCTCCACATCCATCTTCAAGCTACGGGGAACCAA 1983
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 1984 CTTCCCTCCACTAGTGGGAACCTTACATACACTTCTCTC----- 2021
Db 711 rgsSerProValCysMetGluPheGlnTyrGlnAlaMetGlyHisGlyValAlaLeu 731
QY 2022 -----CAGGACTGACAGCTGCTC-----CTCAGCCAGGCCAG 2055
Db 731 InValValArgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2056 TATGATACCCCGAAGCTGGGAGCCAGGCTTACCTGCCCCCAGCAGCAATTTGTTACCAG 2115
Db 751 GlySerGluTrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2116 GTGCCACAGACGACACAAGAAAGTATCAGCAGGAGGAGGATGGGAA 2163
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu 783

RESULT 7

US-08-936-135-14

; Sequence 14, Application US/08936135

; Patent No. 6054293

GENERAL INFORMATION:
 APPLICANT: Tessier-Lavigne, Marc
 APPLICANT: He, Zhigang
 APPLICANT: Chen, Hang
 TITLE OF INVENTION: Semaphorin Receptors
 NUMBER OF SEQUENCES: 26
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
 STREET: 75 DENISE DRIVE
 CITY: HILLSBOROUGH
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94010
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/936,135
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: OSMAN, RICHARD A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: UC97-288-2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 343-4341
 TELEFAX: (650) 343-4342
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 926 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-936-135-14

Alignment Scores:
 Pred. No.: 1.68e-30 Length: 926
 Score: 412.50 Matches: 205
 Percent Similarity: 36.60% Conservative: 109
 Best Local Similarity: 23.89% Mismatches: 334
 Query Match: 10.39% Indels: 210
 DB: 3 Gaps: 37

US-10-060-830-2 (1-2190) x US-08-936-135-14 (1-926)

QY 1 ATGCCTCTGTCCTCCTGCTTACTTGTCTGCTCTGCTGCTC-----GAGGACGCT 54
 DB 1 MetAspMetPheProLeuThrTrpValPheLeuAlaLeuTyrPheSerGlyHisGluVal 20
 QY 55 GGAGCCCAAGGATGGATGGACACACTGTACTAGCCCTGAGAGTGAACCCCTT 114
 DB 21 ArgSerGlnAspProProCysGlyArgProAsnSerLysAspAlaGlyTyrile 40
 QY 115 ACATCCATAAATACCACAGACCTATCCCAACAGCACTGTTTGTGAATGGAGATCCGT 174
 DB 41 ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluTrpIleValTyr 60
 QY 175 GTA---AAGATGGGAGAGAGTTCGCATCAAAATTGCT---GACTTTGACATTGAAGAT 228
 DB 61 AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGluLys 80
 QY 229 TCTGATCTTCTCACTTAATTAATTTAGATTTATATGAATTTGAGTTCAGCAGAACT 288
 DB 81 Hisasp---CysLysTyrAspPheIleGluLeuArgAspGlyAspSerGluSerAlaasp 99
 QY 289 GAAATAGCAAAATACTGTGGTCTGGGGTTCGAATGAAC-----CATTCAAATGAA 339
 DB 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProthriIleIle 114
 QY 340 TCAAAAGCAATGAATACATTCGTTGATGAGTGAATCCATGTTTCTGGAGCCGGA 399

DB 115 SerSerGlySerValLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly 134
 QY 400 TTTTGGCCTCATCTCTGTTATAGATAACAAGATCTAATTACTTGTGTTGGAGACTGCA 459
 DB 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
 QY 460 TCCAATTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGTTGCTGCTCTCCTTTT 519
 DB 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
 QY 520 GCTGAGATATCTGAACAATTCCTCATGGATATAGAGATTCCTCGCATTCGTCAGCGCT 579
 DB 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
 QY 580 GGTGTGCATCGAGGAGTAGTGCAACACAGCTTGGCGGCCAAATCAGT----- 627
 DB 182 LysProArgMetGluIleIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu 201
 QY 628 -----GTTCTAATTAGTAAAGGATTCC----- 651
 DB 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis 221
 QY 652 -----TATTGAAAGTTCTTTGGCTAACACGTCACATCTGTG 690
 DB 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSer 241
 QY 691 GTGGACACTATCTACAAGTCTTTTACATTTAAGCAAGT----- 732
 DB 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
 QY 733 -----GGATGTTAT 741
 DB 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProProGluAsnPheGlnCysAsn 278
 QY 742 GGAACACTGGGATGGAGTCTGTGTGATCGCGATCCTCAATAACAGATCATCTGTG 801
 DB 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer 297
 QY 802 CTGGAGTGGACTGACCACACAGGCGACAGACAGTGGAAACCCCAAAAGCCAGGCTG 861
 DB 298 -----ThrPheSerAspGlyArgTyrTrpThrProGlnGlnSerArgLeu 311
 QY 862 AAAAAACCTGACCGCTTGGGCTGCTTTGCCACTGATGTAATACACAGTGTGTACAAATA 921
 DB 312 HisGlyAspAspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnVal 331
 QY 922 GATTTGAATAAGGAAAAAATAACAGGCATTATAACCACTGGA-----TCCACCATG 975
 DB 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgLeu 351
 QY 976 GTGGAGCACAATTAATCTGTCTGCTACAGAAATCTGTACAGTGTATGATGGCAGAAA 1035
 DB 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
 QY 1036 TGGACTGTGTACAGAGAGCCCTGGTGAGCAAGATAGATATTTCAAGGAAACAAAGAT 1095
 DB 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsp 389
 QY 1096 TATCACCAGGATGTGCGTAATAACTTTTGGCCACCAATTAATTCACCGTATTATTAGAGT 1155
 DB 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgIle 409
 QY 1156 AATCCTACCCCAATGGCAGCAGAAAATGCCATGAATGGAGCTGCTCGGATGTCAGTTT 1215
 DB 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
 QY 1216 ATTCTTAAGGTGCTCTCCCAAACTTACTCAACCTCCACCTCTCGGACACGAATGAC 1275
 DB 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
 QY 1276 CTCRAAAACACTACAGCCCTCCAAAATAGCCAAAGTCGTGCCCAAAATTT----- 1329

Db 751 GlySerGluTrpLysHisGlyArgIleLeuLeuProSerTyrAsp---MetGluTyrGln 769
QY 2116 GTGCCACAGCAGCACACAGAACTATCAGGAGCAGGAGGATGGGAA 2163
Db 770 Ile-----ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu 783
RESULT 9
US-08-936-135-18
; Sequence 18, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936.135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 909 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-936-135-18
Alignment Scores:
Pred. No.: 3 53e-29 Length: 909
Score: 398.50 Matches: 199
Percent Similarity: 36.37% Conservative: 112
Best Local Similarity: 23.27% Mismatches: 340
Query Match: 10.04% Indels: 204
DB: 3 Gaps: 36
US-10-060-830-2 (1-2190) x US-08-936-135-18 (1-909)
QY 1 ATGCTCTCTCTCTCTG-----CTCTTACTGTGCTGCTCTCTGCTGCTGAGGAGCGT 54
Db 1 MetAspMetPheProLeuThrTrpValPheLeuAlaLeuTyrPheSerArgHisGlnVal 20
QY 55 GGAGCCACAGAGTGATGGATGTGGACACACTGACTACTAGGCGCCCTGAGAGTGGAACCCCTT 114
Db 21 Ar-glyGlnProAspProProCysGlyArgLeuAsnSerLysAspAlaGlyTyrIle 40
QY 115 ACATCCATAAATACCCACAGCCTATCCACACACTGTTGTGTAATGGAGATCCGT 174
Db 41 ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluTrpIleValTyr 60
QY 175 GTA---AAGATGGAGAGAGAGTTCGCATCAAAATTTGGT---GACTTTGACATTCAAGAT 228
Db 61 AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGluLys 80

QY 229 TCTGATCTCTCTCACTTTAATTACTTGAATTTATATGAATTTGGAGTCAGCAGAACT 288
Db 81 Hisasp---CysLysTyrAspPheIleGluIleArgAspGlyAspSerCluserAlaasp 99
QY 289 GAAATAGCAAAATACTGTGGTCTGGGTTCGAAATGAAC-----CATTCAATTTGAA 339
Db 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProThrIleIle 114
QY 340 TCAAAGCAATGAATCATCTGCTTTCATGAGTGAATCCATGTTCTTCTGGAGCGGA 399
Db 115 SerSerGlySerMetLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly 134
QY 400 TTTTGGCCTCATCTCTGTTATAGATAAACAAGATCTAATTACTTGTGTTGGCACTGCA 459
Db 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 460 TCCAATTTTGGAACTGAGTTCAGTAAGTACTCCAGCTGCCAGCTGTTGCTGCTCTCTTT 519
Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
QY 520 CTGAGATATCTGGAAACAATTCCTCATGATATAGAGATTCCTGCCATTTGTCATGGCT 579
Db 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 580 GGTGTGCATGAGGAGTAGTGTCAAAACACAGCTTGGCGGCCAAATCAGT----- 627
Db 182 LysProLysMetGluIleIleLeuGlnPheLeuIlePheAspLeuGluHisaspProLeu 201
QY 628 -----GTTGTAATTAGTAAGGTATTCCTCC----- 651
Db 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis 221
QY 652 -----TATTATGAACTTTCTTGGCTAACACACTCAGTCTGTCGTGAGTGG 690
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerGluLeuArgSer 241
QY 691 GTGGACACTTATCTACAAAGCTTTTTCAC----- 720
Db 242 ThrGlyIleLeuSerLeuThrPheHisThrAspMetAlaValAlaLysaspGlyPheSer 261
QY 721 -----TTTAAGACAAGTGGATGTTATGTAACACTG 750
Db 262 AlaArgTyrTyrLeuValHisGlnGluProLeuGluAsnPheGlnCysAsnValProLeu 281
QY 751 GGGATGGAGTCTGTGTGATCCGATCCTCAAAATACAGCATCTCTCTGCTGAGTGG 810
Db 282 GlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer----- 297
QY 811 ACTGACCACAGCGCAGAGAACAGTTCGAAACCCAAAGCCAGGCTGAAACACCT 870
Db 298 -----ThrTyrSerAspGlyArgTrpThrProGlnGlnSerArgLeuHisGlyAsp 314
QY 871 GGACCGCTTGGCTGCTTTTCCACTGATGAATACCACTGCTGTACAAATAGATTGTAAT 930
Db 315 AspAsnGlyTrpThrProAsnLeuAspSerAsnLysGluTyrLeuGlnValaspLeuArg 334
QY 931 AAGGAAAAAGAAATAACAGGCATTATAACCACTGGA-----TCCACCATGTTGGAGCAC 984
Db 335 PheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGluThrGlnAsn 354
QY 985 AATTACTATGTCTGCTGCTACAGATCCTGTACAGATGATGATGGCAGAGAAATGGACTGTG 1044
Db 355 GlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAspTrpMetVal 374
QY 1045 TACAGAGAGCCTGGTGTGGAGCAAGATAGATATTCAAGGAAACAAAGATTATCACCAG 1104
Db 375 TyrArgHisGly-----LysAsnHisLysValPheGlnAlaAsnAspAlaThrGlu 392
QY 1105 GATGTCGCTTAACCTTTTGGCCACCAATTTTGGCCAGCTTTTATTAGATGTAATCCTACC 1164
Db 393 ValValLeuAsnLysLeuHisAlaProLeuLeuThrArgPheValArgIleArgProGln 412

Db 41 ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluTrpIleValTyr 60
QY 175 GTA---AAGATGGAGAGAGCTCGCATCAATTTGT---GACTTTGACATTCAGAT 228
Db 61 AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGlnLys 80
QY 229 TCTGATCTTGTCACTTTAATTTACTTGAAGATTTAATGGAATTTGGAGTGCAGCAGAAT 288
Db 81 HisAsp---CysLysTyrAspPheIleGluIleArgAspGlyAspSerGluSerAlaAsp 99
QY 289 GAAATAGCAATACTGTGCTGGGTGGGTTGCAATGAAC-----CAITCAATTGAA 339
Db 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProThrIleIle 114
QY 340 TCAAAAGCAATGAATACATTCATGCTGTTTCAGTGAATCCATGTTCTCGACGCGGA 399
Db 115 SerSerGlySerMetLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly 134
QY 400 TTTTGGCTCATCTCTCTTATAGATAAACAAGATCTAATTTACTTGTGGGACTGCA 459
Db 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 460 TCCAATTTTGGACCTGAGTTAGTACTGCTGCCAGCTGGTGTCTGCTTCTCTTT 519
Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
QY 520 GCTGAGATCTGGACAATCTCTCATGATATAGAGATTCCTGCCATTTGTCATGCT 579
Db 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 580 GGTGTGCATCGAGGAGTGTGTCACACAGCTTGGCGGCGCAATCAGT----- 627
Db 182 LysProLysMetGluIleLeuGlnPheLeuIlePheAspLeuGluHisAspProLeu 201
QY 628 -----GTTGTAATTAAGTATTCCTCC----- 651
Db 202 GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis 221
QY 652 -----TATTATGAAGTCTTTTGTGCTTAACACCTCACATCTGTG 690
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerGluLeuArgSerSer 241
QY 691 GTGGGACCTTATCTACAAGTCTTTTACA----- 720
Db 242 ThrGlyIleLeuSerLeuThrPheHisThrAspMetAlaValAlaLysAspGlyPheSer 261
QY 721 -----TTTAGACAAGTGGATGTTATGGACACATG 750
Db 262 AlaArgTyrTyrLeuValHisGlnGluProLeuGluAsnPheGlnCysAsnValProLeu 281
QY 751 GGGATGAGTCTGCTGATCGCGATCCTCAATAACAGCATCATCTGCTGGAGTGG 810
Db 282 GlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSer----- 297
QY 811 ACTGACCACAGCGGCAAGACAGTGTGGAACCAAAAGCCAGGCTGAAAAAACCT 870
Db 298 -----ThrTyrSerAspGlyArgTyrThrProGlnGlnSerArgLeuHisGlyAsp 314
QY 871 GGACCCCTTGGCTGCTTTTGCACATGATGAATACCACTGTTTACAATAGATTTGAAT 930
Db 315 AspAsnGlyTrpThrProAsnLeuAspSerAsnLysGluTyrLeuGlnValAspLeuArg 334
QY 931 AGGAAAGAAATACACGCAATTATAACCTGGA-----TCCACCATGTTGGAGCAC 984
Db 335 PheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGluThrGlnAsn 354
QY 985 AATTACTATGTCTGCTCACAGAATCTGTACAGTATGATGGCAGAAATGGACTGTG 1044
Db 355 GlyTyrValLysSerTyrLysLeuGluValSerThrAsnGlyLysIleAspTrpMetVal 374
QY 1045 TACAGAGAGCTGCTGTGGAGCAAGATAAGATTTTCAAGGAAACCAAGATTCACCAAG 1104
Db 375 TyrArgHisGly-----LysAsnHisLysValPheGlnAlaAsnAsnAspAlaThrGlu 392

QY 1105 GATGTGCGTAATAACTTTTGTGCCACCAATTAATTCACGTTTATTATAGATGAATCCTACC 1164
Db 393 ValValLeuAsnLysLeuHisAlaProLeuLeuThrArgPheValArgIleArgProGln 412
QY 1165 CAATGGCAGCAGAAAATGGCATGAAATGGAGCTGCTGGATGTCAGTTTATCTCTAAA 1224
Db 413 ThrTrpHisSerGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1225 GGTGCTCTCCAAAATTTACTCAACCTCCACCTCTCCGGAACGACGAATGACCTCAAAAC 1284
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMetLeuSerGly 443
QY 1285 ACTCAGCCCTCCAAAATTAATGCAAGGCTGCTGCCCAAAATTT----- 1329
Db 444 LeuIleAlaAspSerGlnIleSerAlaSerSerThrGlnGluTyrLeuTrpSerProSer 463
QY 1330 AGCNAACCATACAACTCGCAGTAGCAATGAATTTCTCTGCACACAGACAGAAACAAC 1389
Db 464 AlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgIleProGlnAlaGlnPro 483
QY 1390 GCCAGTCTGATATCAGAAATACTACCTAATCTCAATCTCAATCTCAATCTCAATCTCA 1443
Db 484 GlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGlyValIleIle 503
QY 1444 -----CGCTGGCTGCAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1482
Db 504 GlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPheValArgLys 523
QY 1483 CTCACTACTCTCATCTCATATTTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1527
Db 524 PheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAspProArgThr 543
QY 1528 AACGAAAGAAAACCTGAGAGCACC-----TATGACTTACTTACTTACTTACTTACTTACTT 1581
Db 544 GlnGlnProLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1582 GTTGTGGAAGGAATGAAGCAGTTTCTCTGCTGCAAAAGCAGTGGACCATGAGGAAACC 1641
Db 559 -----AspIleArgArgPheAspPro-----Ile 566
QY 1642 CGAGTTCGTATAGCAGCAGCAGGAAAGTAAATCAGCTGAGTCAAGAGAGATCACCAGATG 1701
Db 567 ProAlaGlnTyrValArgValTyrProGluArgTyrSerProAlaGlyIleGlyMetArg 586
QY 1702 CTGAGGCTGATCTGCAGATAT-----GCTCAGCCTAGTGTAGGAGAAATTTGTTGT 1755
Db 587 LeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValLysThrLeuGlyPro 606
QY 1756 ACACCTTCATCAAGATCTACCTTTAAACCA-----GAAGAAGGAAAGAAAGCA 1803
Db 607 ThrValLysSerGluGluThrThrThrProTyrProThrGluGluAlaThrGluCys 626
QY 1804 GGC-----TATGACAGCCTAGATCTTACAACTCACCAGGCGAGGAG-- 1846
Db 627 GlyGluAsnCysSerPheGluAspAspLysAsp-LeuGlnLeuProSerGlyPheAsnCy 646
QY 1847 -----TTTATCATGCCT-----ATGCTGAACCACTCCCAA---TTACG 1881
Db 646 sAsnPheAspPheLeuGluGluProCysGlyTrpMetTyrAspHisAlaLysTrpLeuArg 666
QY 1882 GGGCTGAGTATGCAACCAATCATCATGAGCATGTCAGGCGACCCCAAC----- 1933
Db 666 g-----ThrThrTrpAlaSerSerSerSerProAsnAspArgTh 679
QY 1933 ----- 1933
Db 679 rPheProAspAspArgAsnPheLeuArgLeuGlnSerAspSerGlnArgGluGlyGlnTy 699
QY 1934 -CTTCAGTTGTCAGCCCTCCACATCTTCAAGGCTACGGGGAACCAACCTCCCCCA 1992
Db 699 rAlaArgLeuIleSerProProValHisLeu-----ProArgSerProV 714

```
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 902 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 952 952 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1094 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1490 1490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1690 1690 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1839 1839 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1997 1997 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2196 2196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 587 592 NFIPLA -> T (IN VARIANT 2).
SQ SEQUENCE 2211 AA; 248981 MW; CBBF90B738667C45 CRC64;

Alignment Scores:
Pred. No.: 6,94e-18 Length: 2211
Score: 361.00 Matches: 79
Percent Similarity: 57.83% Conservatives: 17
Best Local Similarity: 47.59% Mismatches: 60
Query Match: 8.71% Indels: 10
DB: 1 Gaps: 3

US-10-060-830-1 (1-2280) x FA5_BOVIN (1-2211)
QY 799 AAGCAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATCGCGGATCCTCAA 858
Db 2049 GluValAsnGlyCysSerThrProLeuGlyMetGluSerGlyLeuIleGluAsnLysGln 2068
QY 859 ATACACATCATCTGTGCTGGAG-----TGGACTACACACAGGCGGACAGAACAGT 912
Db 2069 IleThrAlaSerPheLysSerTrpTrp-----GlyAsnTyr 2082
QY 913 TGGAAACCCAAAGGACGCTGAAACACCTGGA-----CGCCTTGGGCTGCTTTT 966
Db 2083 TrpGluProPheLeuAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2102
QY 967 GCCACTGATGAATACCACTGCTTACAAATAGATTGTAATAGGAAAAAGAAATAACAGGC 1026
Db 2103 AlaAsnAsnAsnGlnTrpLeuGlnIleAspLeuLeuLysIleLysIleThrAla 2122
QY 1027 ATTATACCACTGGATCCCATCGGTGGAGCAACAATTACTATGTCTGCTGCTACAGATC 1086
Db 2123 IleValThrGlnGlyCysLysSerLeuSerSerGluMetTyrValLysSerTyrThrIle 2142
QY 1087 CTGPACAGTGTATGGCGAAGATGACCTGTGACAGAGAGCCTGGTGTGGAGCAAGAT 1146
Db 2143 HisTyrSerAspGlnGlyThrAspTrpLysProTyrArgGluLysSerSerMetValAsp 2162
QY 1147 AAGTATTTCAAGGAACAAGATTATCACCGAGTGTGCGTAATAACTTTTTCGCCACCA 1206
Db 2163 LysIlePheGluGlyAsnAsnValArgLysHisValLysAsnPheAsnProPro 2182
QY 1207 ATTTATTCACCTTTTATAGTAGTGAATCCCTACCAAGTGGCAGCAAGAAATGCCATGAA 1266
Db 2183 IleIleSerArgPheIleArgIleProLysThrTrpAsnGlnSerIleAlaLeuArg 2202
QY 1267 ATGGAGCTGCTCGGATGT 1284
Db 2203 LeuGluLeuPheGlyCys 2208
```

```
RESULT 11
MFGM_BOVIN STANDARD; PRT; 427 AA.
ID Q95114; Q27959; P79344;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE (MGP57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE SP47) (BP47) (Components 15/16).
GN MFG-E8.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
SEQUENCE FROM N.A. PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC STRAIN=Holstein; TISSUE=Mammary gland;
RX MEDLINE=97008954; PubMed=8856064;
RA Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA Petersen T.E.;
RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT milk fat globules";
RL Eur. J. Biochem. 240:628-636(1996).
RN [2]
SEQUENCE OF 18-427 FROM N.A.
RP TISSUE=Mammary gland;
RX MEDLINE=96125736; PubMed=8541316;
RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA Matsuda T.;
RT "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
RT monoclonal antibodies raised against bovine milk fat globule
RT membrane.";
RL Blochim. Biophys. Acta 1245:385-391(1995).
RN [3]
SEQUENCE OF 19-427 FROM N.A.
RP TISSUE=Testis;
RA Ensslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RP TISSUE=Milk;
RX MEDLINE=93250576; PubMed=8485470;
RA Mather I.H., Banghart L.R., Lane W.S.;
RT "The major fat-globule membrane proteins, bovine components 15/16 and
RT guinea-pig GP 55, are homologous to MGP-E8, a murine glycoprotein
RT containing epidermal growth factor-like and factor V/VIII-like
RT sequences.";
RL Biochem. Mol. Biol. Int. 29:545-554(1993).
CC -1- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC NANNARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
CC LACKS 53 AMINO ACIDS WITHIN THE F578 TYPE C 1 DOMAIN.
CC -1- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
CC -1- PTM: THE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH
CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91895; CAA62997.1; -
DR EMBL; S80643; AAB35894.2; -
```


AC P12259; Q14285;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Coagulation factor V precursor (Activated protein C cofactor).
 GN F5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE-92232668; PubMed-1567832;
 RX Cripe L.D., Moore K.D., Kane W.H.;
 RA "Structure of the gene for human coagulation factor V.";
 RT Biochemistry 31:3777-3785(1992).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE-87260886; PubMed-3110773;
 RX Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
 RA Hewick R.M., Kaufman R.J., Mann K.G.;
 RT "Complete cDNA and derived amino acid sequence of human factor V.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
 RN [3]
 RP SEQUENCE OF 1-1600 FROM N.A.
 RX MEDLINE-88107560; PubMed-2827731;
 RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
 RT "Cloning of cDNAs coding for the heavy chain region and connecting
 region of human factor V, a blood coagulation factor with four types
 of internal repeats.";
 RL Biochemistry 26:6508-6514(1987).
 RN [4]
 RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
 RX MEDLINE-86313665; PubMed-3092220;
 RA Kane W.H., Davie E.W.;
 RT "Cloning of a cDNA coding for human factor V, a blood coagulation
 factor homologous to factor VIII and ceruloplasmin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
 RN [5]
 RP PARTIAL SEQUENCE FROM N.A.
 RC TISSUE-Fibroblast;
 RX MEDLINE-93203619; PubMed-8454869;
 RA Shen N.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,
 RA Edgington T.S.;
 RT "The serine protease cofactor factor V is synthesized by
 lymphocytes.";
 RL J. Immunol. 150:2992-3001(1993).
 RN [6]
 RP SULFATION.
 RX MEDLINE-94264012; PubMed-8204629;
 RA Pittman D.D., Tomkinson K.N., Michnick D., Seligsohn U.,
 RA Kaufman R.J.;
 RT "Posttranslational sulfation of factor V is required for efficient
 thrombin cleavage and activation and for full procoagulant activity.";
 RL Biochemistry 33:6952-6959(1994).
 RN [7]
 RP SULFATION.
 RX MEDLINE-90366699; PubMed-2168225;
 RA Hortic G.L.;
 RT "Sulfation of tyrosine residues in coagulation factor V.";
 RL Blood 76:946-952(1990).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
 RX MEDLINE-20052169; PubMed-10586886;
 RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
 RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
 RA Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 coagulation factor V.";
 RL Nature 402:434-439(1999).
 RN [9]
 RP VARIANT APCR GLN-534.
 RX MEDLINE-94217810; PubMed-8164741;

RA Bertina R.M., Koelman B.P.C., Koster T., Rosendaal F.R.,
 RA Dirven R.J., de Ronde H., van der Velde P.A., Reitsma P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 activated protein C.";
 RL Nature 369:84-87(1994).
 RN [10]
 RP VARIANT APCR GLY-334, AND VARIANT LYS-513.
 RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
 RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese.";
 RL Blood 91:1135-1139(1998).
 RN [11]
 RP VARIANT APCR THR-334.
 RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
 RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
 resistance to activated protein C.";
 RL Blood 91:1140-1144(1998).
 CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
 chain, noncovalently bound. The interaction between the two chains
 is calcium-dependent.
 CC -1- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
 REPEATS.
 CC -1- PTM: Thrombin activates factor V proteolytically to the active
 cofactor, factor Va (formation of a heavy chain at the N-
 terminus and a light chain at the C-terminus).
 CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
 ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
 CC -1- DISEASE: OMEN PARAHOMOPHILIA, AN HEMORRHAGIC DIATHESIS, IS DUE
 TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
 FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APCR
 MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
 A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
 IMPLANTATION.
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
 2 PLASTOCYANIN-LIKE REPEATS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: STRONG TO COAGULATION FACTOR VIII.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 DR EMBL; L32779; AAB59401.1; JOINED.
 DR EMBL; L32755; AAB59401.1; JOINED.
 DR EMBL; L32756; AAB59401.1; JOINED.
 DR EMBL; L32757; AAB59401.1; JOINED.
 DR EMBL; L32758; AAB59401.1; JOINED.
 DR EMBL; L32759; AAB59401.1; JOINED.
 DR EMBL; L32760; AAB59401.1; JOINED.
 DR EMBL; L32761; AAB59401.1; JOINED.
 DR EMBL; L32762; AAB59401.1; JOINED.
 DR EMBL; L32763; AAB59401.1; JOINED.
 DR EMBL; L32764; AAB59401.1; JOINED.
 DR EMBL; L32765; AAB59401.1; JOINED.
 DR EMBL; L32766; AAB59401.1; JOINED.
 DR EMBL; L32767; AAB59401.1; JOINED.
 DR EMBL; L32768; AAB59401.1; JOINED.
 DR EMBL; L32769; AAB59401.1; JOINED.
 DR EMBL; L32770; AAB59401.1; JOINED.
 DR EMBL; L32771; AAB59401.1; JOINED.
 DR EMBL; L32772; AAB59401.1; JOINED.
 DR EMBL; L32773; AAB59401.1; JOINED.
 DR EMBL; L32774; AAB59401.1; JOINED.
 DR EMBL; L32775; AAB59401.1; JOINED.
 DR EMBL; L32776; AAB59401.1; JOINED.
 DR EMBL; L32777; AAB59401.1; JOINED.
 DR EMBL; L32778; AAB59401.1; JOINED.
 DR EMBL; M16967; AAB52424.1;

DR EMBL; M14335; AAB59532.1; -;
 DR PIR; A25897; A25897.
 DR PIR; A28028; A28028.
 DR PDB; 1CZS; 26-NOV-99.
 DR PDB; 1CZT; 26-NOV-99.
 DR PDB; 1CZV; 26-NOV-99.
 DR Genew; HGNC:3542; F5.
 DR MIM; 227400; -;
 DR MIM; 134400; -;
 DR MIM; 188055; -;
 DR MIM; 227310; -;
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR000421; FA5H_C.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
 DR PROSITE; PS01285; FA58C.1; 2.
 DR PROSITE; PS01286; FA58C.2; 2.
 KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
 KW Repeat; Polymorphism; Disease mutation; Thrombophilia; 3D-structure.
 FT SIGNAL 1 28
 FT CHAIN 29 2224 COAGULATION FACTOR V.
 FT CHAIN 29 737 HEAVY CHAIN.
 FT PEPTIDE 738 1573 ACTIVATION PEPTIDE (CONNECTING REGION).
 FT CHAIN 1574 2224 LIGHT CHAIN.
 FT DOMAIN 30 329 F5/8 TYPE A 1.
 FT DOMAIN 30 329 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 203 329 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 348 684 F5/8 TYPE A 2.
 FT DOMAIN 348 526 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 336 684 PLASTOCYANIN-LIKE 4.
 FT DOMAIN 692 1573 B.
 FT DOMAIN 895 928 2 X 17 AA TANDEM REPEATS.
 FT REPEAT 895 911 1-1.
 FT REPEAT 912 928 1-2.
 FT SIMILAR 1135 1148 TO 14 AA REPEATS IN BOVINE FA5.
 FT DOMAIN 1185 1501 [TNP]-L-S-P-D-L-S-Q-T.
 FT REPEAT 1185 1193 2-1.
 FT REPEAT 1194 1202 2-2.
 FT REPEAT 1203 1211 2-3.
 FT REPEAT 1212 1220 2-4.
 FT REPEAT 1221 1229 2-5.
 FT REPEAT 1230 1238 2-6.
 FT REPEAT 1239 1247 2-7.
 FT REPEAT 1248 1256 2-8.
 FT REPEAT 1257 1265 2-9.
 FT REPEAT 1266 1274 2-10.
 FT REPEAT 1275 1283 2-11.
 FT REPEAT 1284 1292 2-12.
 FT REPEAT 1293 1301 2-13.
 FT REPEAT 1302 1310 2-14.
 Alignment Scores:
 Pred. No.: 6,83e-17 Length: 2224
 Score: 347.00 Matches: 76
 Percent Similarity: 58.43% Conservativity: 21
 Best Local Similarity: 45.78% Mismatches: 59
 Query Match: 8.38% Indels: 10
 DB: 1 Gaps: 3
 US-10-060-830-1 (1-2280) x FA5_HUMAN (1-2224)
 QY 799 AGACAGTGGATGTATGGAACATCGGGATGGAGTCTGGTGTGATCGCGATCCTCAA 858
 Db 2062 GluValAsnGlyCysSerThrProLeuGlyMetGluAsnGlyLysIleGluAsnLysGln 2081
 QY 859 ATAACAGCATCTGTGCTGCAG-----TGGACTGACCACACAGCGCAGAGAACAGT 912
 Db 2082 IleThrAlaSerSerPheLysLysSerThrPrpGlyAspTyr----- 2095
 QY 913 TGAACACCCAAAAAGCCAGCGTGAAGAAACCTGGA-----CCGCGCTGGGCTGCTTTT 966

Db 2096 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2115
 QY 967 GCCACTGATGAATACCACTGGTTACAAATAGATTTTGAATAAGGAAAGAAAATAACAGGC 1026
 Db 2116 AlaAsnAsnLysGlnTrpLeuGluIleAspLeuLeuLysIleLysLysIleThrAla 2135
 QY 1027 ATTATAACCACTGGATCCACCATGGTGGAGCACAAATTACTATGTCTGTCCTACAGATC 1086
 Db 2136 IleIleThrGlnGlyCysLysSerLeuSerGluMetTyrValLysSerTyrThrIle 2155
 QY 1087 CTGTACAGTGTATGGGCGAGAAATGGACTGTGTACAGAGAGCGCTGGTGGAGCAAGAT 1146
 Db 2156 HisTyrSerGluGlnGlyValGluTrpLysProTyrArgLeuLysSerSerMetValasp 2175
 QY 1147 AAGATATTTCAAGGAACAAAGATTATCACCAGGATGCGGTAATTAACATTTTGGCCACCA 1206
 Db 2176 LysIlePheGluGlyAsnThrAsnThrLysGlyHisValLysAsnPheAsnProPro 2195
 QY 1207 ATTATTGACGCTTTATTAGAGTGAATCCTACCCCAATGGCAGCAGCAAAATTCCTCATGAA 1266
 Db 2196 IleIleSerArgPheIleArgValIleProLysThrTrpAsnGlnSerIleThrLeuArg 2215
 QY 1267 ATGGAGCTGCTCGGATGT 1284
 Db 2216 LeuGluLeuPheGlyCys 2221
 RESULT 13
 MFGM_HUMAN STANDARD; PRT; 387 AA.
 AC Q08431;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-EB) (HMFG)
 DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
 GN MFGEB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Breast, and Breast carcinoma;
 RA MEDLINE=96213908; PubMed=8639264;
 RX Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
 RT "Cloning and sequence analysis of human breast epithelial antigen
 BA46 reveals an RGD cell adhesion sequence presented on an epidermal
 growth factor-like domain.";
 RL DNA Cell Biol. 15:281-286(1996).
 RN [2]
 RP SEQUENCE OF 170-387 FROM N.A.
 RC TISSUE-Mammary gland;
 RX MEDLINE=91371351; PubMed=1909932;
 RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
 RA Ceriani R.L.;
 RT "A Mr 46,000 human milk fat globule protein that is highly expressed
 in human breast tumors contains factor VIII-like domains.";
 RL Cancer Res. 51:4994-4998(1991).
 RN [3]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RC TISSUE-Milk;
 RX MEDLINE=98194924; PubMed=9535276;
 RA Gluffida M.G., Cavaletto M., Giunta C., Conti A.,
 RA Godovac-Zimmermann J.;
 RT "Isolation and characterization of full and truncated forms of human
 breast carcinoma protein BA46 from human milk fat globule membranes.";
 RL J. Protein Chem. 17:143-148(1998).
 RN [4]
 RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
 RX MEDLINE=99342076; PubMed=10411933;
 RA Haeggqvist B., Naeslund J., Sletten K., Westermark G.T., Mucchiano G.,
 RA Tjernberg L.O., Nordstedt C., Engstrom U., Westermark P.;

RT "Medin: an integral fragment of aortic smooth muscle cell-produced
 RL lactadherin forms the most common human amyloid";
 RT Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE:97405885; PubMed:9260929;
 RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
 RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
 RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
 RT (RGD)-dependent cell adhesion";
 RL DNA Cell Biol. 16:861-869(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 CC SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
 CC MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
 CC -1- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
 CC AMINO ACID 264 AND 273.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U59516; AAC50949.1; -;
 DR EMBL: S56151; AAB19771.1; -;
 DR HSP: P08709; 1BF9.
 DR Genew: HGNC:7036; MFE88.
 DR MM: 602281;
 DR SMART: SM00181; EGF_1;
 DR SMART: SM00231; FA58C; 2.
 DR PROSITE: PS00022; EGF_1;
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01285; FA58C_1; 2.
 DR PROSITE: PS01286; FA58C_2; 2.
 KW Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
 FT SIGNAL 1 23
 FT CHAIN 24 387
 FT CHAIN 202 387
 FT CHAIN 268 317
 FT DOMAIN 24 67
 FT DOMAIN 70 225
 FT DOMAIN 230 387
 FT SITE 46 48
 FT SITE 27 38
 FT DISULFID 32 55
 FT DISULFID 57 66
 FT DISULFID 70 225
 FT DISULFID 212 387
 FT DISULFID 230 387
 FT CARBOHYD 238 325
 FT CARBOHYD 325 325
 FT CARBOHYD 329 329
 FT CARBOHYD 350 350
 SQ SEQUENCE 387 AA; 43123 MW; 2EE6571DEC83782D CRC64;
 Alignment Scores:
 Pred. No.: 9.26e-17 Length: 387
 Score: 343.00 Matches: 71
 Percent Similarity: 57.23% Conservative: 28
 Best Local Similarity: 41.04% Mismatches: 68
 Query Match: 8.28% Indels: 6
 DB: 1 Gaps: 3

US-10-060-830-1 (1-2280) x MFGM_HUMAN (1-387)
 QY 775 TTATCTACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACACTGGGATGGAG 834
 Db 218 LeuArgPheGluLeuLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 237
 QY 835 TCTGGTGTGATCGCGGATCTCTCAATAACACAGCATCATCTGCTGGAG---TGGACTGAC 891
 Db 238 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerTyrLysThrTTPGlyLeu 257
 QY 892 CACACAGGCAAGAGAAGACAGTGGAAACCCAAAAAACCAGGCTGAAAAACCTGGA--- 948
 Db 258 HisLeu-----PheSerTTPAsnProSerTyrAlaArgLeuAspLysGlnGlyAsn 274
 QY 949 ---CCGCTTGGGCTGCTTTTGGCCACTGATCAATACAGTGGTTTACAAATAGATTTCGAAT 1005
 Db 275 PheAsnAlaTrpValAlaGlySerTyrGlyAsnAspGlnTrpLeuGlnValAspLeuGly 294
 QY 1006 AAGGAAAGAAATACAGGCAATATATACACACTGGATCCACCATGGTGGAGCACAATAC 1065
 Db 295 SerSerLysGluValThrGlyIleIleThrGlnGlyAlaArgAsnPheGlySerValGln 314
 QY 1066 TATGTGCTGCTCAGAAATCCTCTACAGTATGATGGCAGAAATGGACTGTGTACAGA 1125
 Db 315 PheValAlaSerTyrLysValAlaTyrSerAsnAspSerAlaAsnTrpThrGluTyrGln 334
 QY 1126 GAGCTGTGTGGAGCAAGATAAGATATTTCAAGGAAACAAAGATATATCACCAGGATGTG 1185
 Db 335 AspProArgThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisSerHisLys 354
 QY 1186 CGTAATAACTTTTCCACCAATATTTGCAGCTTTTATTAGTAGTCAATCTACCAATGG 1245
 Db 355 LysAsnLeuPheGluThrProIleLeuAlaArgTyrValArgIleLeuProValAlaTirp 374
 QY 1246 CAGCAGAAATTTGCCATGAAATGAGCTGCTCGGATGT 1284
 Db 375 HisAsnArgIleAlaLeuArgLeuGluLeuLeuGlyCys 387
 RESULT 14
 MFGM_PIG
 ID MFGM_PIG STANDARD; PRT; 409 AA.
 AC P79385;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Lactadherin (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM) (Sperm
 DE surface protein SP47) (PP47).
 DE MFE8.
 GN Sus scrofa (Pig).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Ensslin M.A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
 CC BINDING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
 CC SPERMATOZOAN.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

Db 388 evalArgilleuProValAlaTrpHisAsnArgIleThrIeuArgValGluLeuIeuG1 408

Qy 1281 ATGT 1284
||||

Db 408 yCys 409

RESULT 15

MFGM_RAT	STANDARD;	PRT;	427 AA.
AC P70490;			
DT 01-NOV-1997 (Rel. 35, Created)			
DT 01-NOV-1997 (Rel. 35, Last sequence update)			
DT 30-MAY-2000 (Rel. 39, Last annotation update)			
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-			
DE acetyl G33 ganglioside synthase) (AGS) (MFGM).			
GN MFC88 OR AGS.			
OS Rattus norvegicus (Rat).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX NCBI_TaxID=10116;			
[1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=Brain;			
RA MEDLINE=96374422; PubMed=8780713;			
RG Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;			
RT "Cloning and expression of cDNA for O-acetylation of GD3			
ganglioside.";			
RL Biochem. Biophys. Res. Commun. 225:932-938(1996).			
CC -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO			
CC PARTICIPATE IN THE O-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.			
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.			
CC -!- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.			
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.			
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.			
CC			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC or send an email to license@isb-sib.ch).			
CC			
DR EMBL; D84068; BAA12210.1; "			
DR HSSP; P00740; 1EDM.			
DR InterPro; IPR000561; EGF-like.			
DR InterPro; IPR001438; EGF II.			
DR InterPro; IPR000421; FA58_C.			
DR Pfam; PF00006; EGF; 2.			
DR Pfam; PF00754; F5_F8_type_C; 2.			
DR PRINTS; PR00010; EGFBL00D.			
DR SMART; SM00181; EGF; 2.			
DR SMART; SM00231; FA58C; 2.			
DR PROSITE; PS00022; EGF_1; 2.			
DR PROSITE; PS01186; EGF_2; 2.			
DR PROSITE; PS01285; FA58C_1; 2.			
DR PROSITE; PS01286; FA58C_2; 2.			
KW Signal; Glycoprotein; Repeat; EGF-like domain; Milk.			
FT SIGNAL	1 22		
FT CHAIN	23 427		
FT DOMAIN	24 61		
FT DOMAIN	64 108		
FT DOMAIN	111 267		
FT DOMAIN	272 427		
FT DISULFID	28 39		
FT DISULFID	33 49		
FT DISULFID	51 60		
FT DISULFID	68 79		
FT DISULFID	73 96		
FT DISULFID	98 107		
FT DISULFID	111 267		
FT DISULFID	234 258		
FT DISULFID	272 427		


```
DR ENBL; AF387547; AAL30178.1; -.
DR InterPro; IPR000859; FA5B_domain
DR InterPro; IPR000431; FA5B_C
DR InterPro; IPR000403; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00734; F5-F8_typeC; 1.
DR PROSITE; PS01180; CUB; 1.
DR CDD; 1180; CUB; 1.
KW DNA-binding; Transcription regul
```

Alignment Scores:			
Pred. No.:	0	Length:	775
Score:	4017.00	Matches:	754
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	96.96%	Indels:	0
DB:	4	Gaps:	0

US-10-060-830-1 (1-2280) x Q96PD2 (1-775)

QY	1	GCGCCGCCCGCCCTGGGCGCGCTCC	60
Db	22	AlaAlaLaProAlaTApAlaLeuProLeuSerArgSerLeuProCysSerAsn	41
QY	61	TCCTCCTCCCTCCATGCCTCTGTTCCTCCTCCTCTACTTGCTGCTCCTGCTGCTGCTC	120
Db	42	SerSerSerPheSerMetProLeuPheLeuLeuLeuLeuValLeuLeuLeuLeuLeu	61
QY	121	GAGCACCTCGAGCCACGAAGAAGTGATGGATGTGGACACACTGTACTAGCCCTGAGACT	180
Db	62	GluAspLagLyAlaGlnGlnGlyAspGlycysGlyHisThrValLeuLeuGlyProGluSer	81
QY	181	GGAAACCCTTACATCCATAAATACTACCCACAGACCATATCCCACAGCACTGTTTGTGAATGG	240
Db	82	GlyThrLeuThrSerIleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrp	101
QY	241	GAGATCCGCTAAAGATGGGAGAGAGATTTCGCATCAAATTTGGTGACTTTGACATTGAA	300
Db	102	GluIleArgVallysMetGlyGluArgValArgIleLysPheGlyAspPheAspIleGlu	121
QY	301	GAITCTGATCTTGTACATTAACTTACTTGAGAAATTTATAATGGAAATGGAGTCACGACA	360
Db	122	AspSerAspSerCysHisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArg	141
QY	361	ACTGAATAGCCAAAATACTGTGCTCGGGTTCGAATCCATGTTCTGTGGCGCGGATTTTGG	420
Db	142	ThrGluIleGlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLys	161
QY	421	GGCAATGAATCATCTGTTCTCATGAGTGAATCCATGTTCTGTGGCGCGGATTTTGG	480
Db	162	GlyAsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeu	181
QY	481	GCCTCATCTCTGTTATAGATAAACAAAGATCTAAATTACTTGTGTGGACACTGCATCCCAAT	540
Db	182	AlaSerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsn	201
QY	541	TTTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGGTTGTCTGCTTCCTTTTCTGTAG	600
Db	202	PheLeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeuPropheAlaGlu	221
QY	601	ATATCTGGACAATTCCTCATGGATATACAGATTCTCCCAATTGTGCATGCTGGTGTG	660
Db	222	IleSerGlyThrIleProHisGlyTyrArgAspSerSerProLeuCysMetaIaGlyVal	241
QY	661	CATGCAAGGAGTAGTGTCAACACGTTGGCGGCCCAATCAGTGTGTATAGTAAGGT	720
Db	242	HisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValIleSerLysGly	261
QY	721	ATTCCCATATTATCAAAAGTCTTTTGGCTAAACAGCTCACATCTGTGTGGGACACATTACT	780
Db	262	IleProTyrTyrgluSerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSer	281
QY	781	ACAAGTCTTTTTACATTTTAAGACAAGTGGATGTTATATGGAACACTGGGGATGAGTCTGCT	840

Db	282	ThrSerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGly	301
Qy	841	GTGATCGCGGATCCTCAATAACAGCATCATCTGTGCTGGAGTGGACTGACACACAGGG	900
Db	302	ValIleAlaAspProGlnIleThrAlaSerSerValLeuGluThrPthrAspHisThrGly	321
Qy	901	CAAGAGAACACTCGAAACCCAAAAAGCCAGGCTGAAAAAACCTCGACCGCCCTCGGCT	960
Db	322	GlnGluAsnSerTrpLysProLysCysAlaArgLeuLysLysProGlyProTrpPala	341
Qy	961	GCTTTTGCCACTGATGAATACCACTGGTTCAAAATAGATTGAATAGGAAAGAAAAATA	1021
Db	342	AlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIle	361
Qy	1021	ACAGCATTAATAACCACTGGATCCACCATGGTGAGCACAAATTACTATGTGTGCTGCTAC	1081
Db	362	ThrGlyIleIleThrThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyr	381
Qy	1081	AGAATCTGTACACTGATGGGCAGAAATGGACTGTGTACAGAGAGCTGTGTGGAG	1141
Db	382	ArgIleLeuTyrSerAspAspGlyGlnLysTrpThrValTyrArgGluProGlyValGlu	401
Qy	1141	CAAGATAAGATATTCAAGGAAACAAAGATTATCACAGGATGTGCGTAACTTTTGTG	1201
Db	402	GlnAspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeu	421
Qy	1201	CCACCAATTATGCACGCTTTATAGAGTGAATCCTACCACCAATGCGCAGCAAAATTGCC	1261
Db	422	ProProlIleAlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAla	441
Qy	1261	ATGAAATGGAGCTGCTCGGATGTCACTTTATTCCTAAAGTTCGCTCCTCCAAACTTACT	1321
Db	442	MetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThr	461
Qy	1321	CAACTCCACTCTCTCGAACAGCAATGACCTCAAAGAACACTACAGCCCTCCAAAAATA	1381
Db	462	GlnProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIle	481
Qy	1381	GCCAAAGTCTGCTGCCCAAAATTTACGCACCACTACAACTCGCAGTAGCAATGATTT	1441
Db	482	AlaLysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPhe	501
Qy	1441	CCTGCACAGACAGAACAAACAACTGCCAGTCCCTGATACAGAATACTACCCGTAACTCCA	1501
Db	502	ProAlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThrPro	521
Qy	1501	AATGTAAACCAAGATGTAGCGCTGGCTTCAGTTCTTGTCCCTGTGCTGTGTCATGTCCTC	1561
Db	522	AsnValThrLysAspValAlaLeuAlaAlaValLeuValProValLeuValMetValLeu	541
Qy	1561	ACTACTCTCATCTCATATTAGTGTGTGCTTGGCAGTGGAGAACAGAAAAAAACT	1621
Db	542	ThrThrLeuIleLeuIleLeuValCysAlaThrPheIleTrpArgAsnArgLysLysLysThr	561
Qy	1621	GAAGCACCTATGACTTACCTTACTGGGACCGGAGGTGTGTGGAAAGGAATGAAGCAG	1681
Db	562	GlulGlyThrTyrAspLeuProTyrTrpAspArgAlaGlyTrpIrpLysGlyMetLysGln	581
Qy	1681	TTTCTTCTCTGCAAGACAGTGGACCATGAGGAAACCCCACTTCGCTATAGCAGCAGCGAA	1741
Db	582	PheLeuProAlaLysAlaValAspHisGluThrProValArgTyrSerSerSerGlu	601
Qy	1741	GTTAATCACCTGAGTCCAGAGAGTCAACACAGTGTGTGAGGCTGACTCTGCAGAGTAT	1801
Db	602	ValAsnHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyr	621
Qy	1801	GCTCAGCCACTGTTAGGAGAAATTGTTGGTACACTTCATCAAGATGCTACCTTTAAACCA	1861
Db	622	AlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysPro	641
Qy	1861	GAAGAGGAAAGAAAGCAGGCTATGCAGACCTAGATCCCTTACAACTCACCGGCGAGGAA	1921

```

Db 642 GluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGlu 661
QY 1921 GTTATCATGCTATGCTGAACACATCCCAATACAGGGGCTGAGTATGCAACCCCAATC 1980
Db 662 ValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIle 681
QY 1981 ATCATGGACATGTCAGGGCACCCACAACTTCAGTTGGTGGCCCTCCACATCCACTTTC 2040
Db 682 IleMetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPhe 701
QY 2041 AAGCTACGGGGAAACCAACCTCCCACTAGTGGGAACCTTACATACACTTCTCCAGG 2100
Db 702 LysAlaThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArg 721
QY 2101 ACTGACAGCTGCTCCTCAGCCAGGCCACTATGATATACCCGGAAGCTGGGAAGCCAGGT 2160
Db 722 ThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGly 741
QY 2161 CTACCTGCCCCAGCAAGTAATGGTGTACCAGGTGCCACAGACACACAAAGATATCAGGA 2220
Db 742 LeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGly 761
QY 2221 GCAGGAGGATGGGAATGTGATGTTTTAAAGAAATCCTT 2262
Db 762 AlaGlyArgAspGlyGluCysAspValPheLysGluIleLeu 775

RESULT 2
Q8TDX2 PRELIMINARY; PRT; 775 AA.
AC Q8TDX2;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE CLCP1.
GN CLCP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Koshikawa K., Osada H., Kozaki K., Konishi H., Masuda A.,
RA Tatenatsu Y., Mitsudomi T., Nakao A., Takahashi T.;
RT "Significant up-regulation of a novel gene, CLCP1, in a highly
RT metastatic lung cancer subline as well as in lung cancers in vivo.";
RL Oncogene 0:0-0(2002)
DR EMBL; AB073146; BAB91138.1; -.
SQ SEQUENCE 775 AA; 85073 MW; 6D07223B21BE5A42 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 775
Score: 4013.00 Matches: 753
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.86% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-1 (1-2280) x Q8TDX2 (1-775)
QY 4 GCCGCCCGCGCTGGCGCGCTCCCTCTCCCGCTCCCTCCCTCCCTCCCTCCCAATCC 63
Db 23 AlaAlaProAlaTyrAlaAlaLeuProLeuSerArgSerLeuProProCysSerAsnSer 42
QY 64 TCCCTCTCCATGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCG 123
Db 43 SerSerPheSerMetProLeuPheLeuLeuLeuLeuLeuValLeuLeuLeuLeuGlu 62
QY 124 GACCTGGAGCCAGCAGGATGGATGGACACACTGACTAGGCCCTGAGAGTGGGA 183
Db 63 AspAlaGlyAlaGlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGly 82
QY 184 ACCCTTACATCCATAAACTACCCACAGACCTATCCCAACAGCACTGTTTGTGAATGGGAG 243

```

```

Db 83 ThrLeuThrSerIleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGlu 102
QY 244 ATCCGTGTAAGATGGGAGAGAGATTCGCATCAAAATTTGGTGTACTTGGACATTGAAGAT 303
Db 103 IleArgValLysMetGlyGluArgValArgLleLysPheGlyAspPheAspIleGluAsp 122
QY 304 TCTGATTCTTGTCACTTTAATTAATCTGAGAATTTAATGGAATTTGGAGTCAGCAGAACT 363
Db 123 SerAspSerCysHisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThr 142
QY 364 GAATAGGCAAAATCTGCTGGTTCGCAATTCGAATTCGAATTCGAATTCGAATTCGAATTC 423
Db 143 GluIleGlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGly 162
QY 424 AATGAAATCACATTCGCTTCATGAGTGGAAATCCATCTTCTGGACGCGGATTTTGGCC 483
Db 163 AsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAla 182
QY 484 TCATACTCTGTATAGATAAACAAGATCTAATTAATTAATTTGGACACTGCATCCAAATTT 543
Db 183 SerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPhe 202
QY 544 TTGAACTGAGTTGAGTAACTGCTGCTGGTTCGCAATTCGAATTCGAATTCGAATTCGAAT 603
Db 203 LeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIle 222
QY 604 TCTGGAACAATTCCTCATGATATAGAGATTCCTCGCATTCATGTCATGGCTGGTGTGCAT 663
Db 223 SerGlyThrIleProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHis 242
QY 664 GCAGGAGTAGTGTCAAAACACGCTGGCGGCCAAATCAGTGTGTAATAGTAAGAGTATT 723
Db 243 AlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValIleSerLysGlyIle 262
QY 724 CCCTATTATGAAGTCTTCTGGCTAAACGTCACATCTGTGGTGGGACACTTATCTACA 783
Db 263 ProTyrTyrGluSerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThr 282
QY 784 AGTCTTTTACATTTAAGACAAGTGGATGTTATGACACACTGGGATGGAGTCTGGTGTG 843
Db 283 SerLeuPheThrPheLeuLysThrSerGlyCystyrGlyThrLeuGlyMetGluSerGlyVal 302
QY 844 ATCCGCGATCTCAAAATCAACAGCATCATCTGTGTGGAGTGGAGTACACACAGGCGAA 903
Db 303 IleAlaAspProGlnIleThrAlaSerSerValLeuGluTrpThrAspHisThrGlyGln 322
QY 904 GAGACAGTTGGAAACCCCAAAAGCCGCTGAAAACCTGAAAACCTGCGCCCTGGCTGCT 963
Db 323 GluAsnSerTrpLysProLysAlaArgLeuLysLysProGlyProProTrpAlaAla 342
QY 964 TTTGCCACTGATGAATACCACTGGTTACAAATAGATTTGAATAAGGAAAAAGAAATAACA 1023
Db 343 PheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysIleThr 362
QY 1024 GGCATTATACCACTGGATCCACCATGCTGGAGCACAATTAATCTATGCTGTGCTGTACAGA 1083
Db 363 GlyIleIleThrThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArg 382
QY 1084 ATCTGTACATGATGATGGCGCAAAATGGACTGTGTACAGAGCCCTGGTGTGGAGCAA 1143
Db 383 IleLeuTyrSerAspAspGlyGlnLysTrpThrValTyrArgGluProGlyValGluGln 402
QY 1144 GATAGATATTTCAAGGAAACAAAGATTTATCACCAGGATGTCGCTAAATCTTTTGGCA 1203
Db 403 AspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuPro 422
QY 1204 CCAATATTGTCAGCTTTTATTAGAGTGAATCCTACCCCAATGCCAGCAGAAAATGGCCATG 1263
Db 423 ProIleIleAlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMet 442
QY 1264 AAAATGGAGCTGCTGGGATGTCAATTTATCTTAAGTCTGCTCTCCAAAACCTACTCAA 1323

```


Db	423	LyshHisSerHisSerGlyGlyPheProAlaThrGlyAlaThrGlnValGluSer	442		
QY	2092	CTCTCCAGGACTGACAGCTGCTCTCAGCCAGGCCAGTATGATACCCGAAAGCTGGG	2151		
Db	443	TyrGlnArgProAlaSerProLysProValGlyGlyTyrAspLysPro--AlaAla	461		
QY	2152	AAGCCAGGCTTACTCTGCCAGACAGCAATTTGGTGATCCAGGTGCCACAGAGCACACAGAA	2211		
Db	462	SerSerPheLeuAspSerArgAsp-----ProAlaSerGlnSerGln	475		
QY	2212	GTATCAGACAGCAAGGATGGG	2235		
Db	476	MetThrSerGlyGlyAspAspGly	483		
RESULT 8					
ID	Q9D696	PRELIMINARY;	PRT;	460 AA.	
AC	Q9D696;				
DT	01-JUN-2001	(T-EMBLrel. 17, Created)			
DT	01-JUN-2001	(T-EMBLrel. 17, Last sequence update)			
DT	01-JUN-2002	(T-EMBLrel. 21, Last annotation update)			
DE	4631413K11R1K	protein.			
GN	4631413K11R1K				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=SKIN;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,				
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,				
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,				
RA	Hayashizaki Y.;				
RT	"Functional annotation of a full-length mouse cDNA collection.";				
RL	Nature 409:685-690(2001).				
CC	-1- SIMILARITY: CONTAINS 1 CUB DOMAIN.				
DR	EMBL; AK014521; BAB29409.1; -				
DR	MGI; MGI:1913936; 4631413K11R1K.				
DR	InterPro; IPR000859; CUB_domain.				
DR	InterPro; IPR004043; LCCL_dom.				
DR	Pfam; PF00431; CUB; 1.				
DR	SMART; SM00042; CUB; 1.				
DR	PROSITE; PS01180; CUB; 1.				
SQ	SEQUENCE 460 AA; 50334 MW; 3CF3356F70A88B93 CRC64;				
Alignment Scores:					
Pred. No.:	2.08e-34	Length:	460		
Score:	507.50	Matches:	169		
Percent Similarity:	35.11%	Conservative:	81		
Best Local Similarity:	23.74%	Mismatches:	161		
Query Match:	12.25%	Indels:	301		
DB:	11	Gaps:	21		
US-10-060-830-1 (1-2280) x Q9D696 (1-460)					
QY	187	CTTACATCCATAAATACCACAGAGCTATCCCAACAGCAGCTCTTTGTGAATGGGAGATC	246		

Db	1	MetThrSerLysAsnTyrProGlyThrTyrProAsnTyrThrValCysGluLysIlelle	20		
QY	247	CGGTAAAGATGGAGAGAGAGTTCGCATCAAAATTTTGGTACTTTGACATTTAGAGATTCT	305		
Db	21	ThrValProLysGlyLysArgLeuLeuArgLeuGlyAspLeuAsnIleGlu--Ser	39		
QY	307	GATCTTGTCACATTAAATTACTTACAGAAATTAATGAATGGATTGGAGTCACAGAGACTGAA	366		
Db	40	LysThrCysAlaSerAspTyrLeuLeuPheSerSer-----AlaThrAspGln	55		
QY	367	ATAGGCAATACTGCTGCTGGGTTGCCAAATGAACCATCAATTGAATCAAAAGGCAAT	426		
Db	56	TyrGlyProTyrCysGly---SerTrpAlaValProLysGluLeuArgLeuAsnSerAsn	74		
QY	427	GAATCACATTGCTGTTTCATGAGTGGAAATCCATGTTTCTGGACGCGGATTTTGGCCCTCA	486		
Db	75	GluValThrValLeuPheLysSerGlySerHisIleSerGlyArgGlyPheLeuLeuThr	94		
QY	487	TACTCTGTTATAGATAACAAGATCTAATTACTTGTGTTGGACACTGCATCAATTTTGG	546		
Db	95	TyrAlaSerSerAspHisProAspLeuIleThrCysLeuGluArgGlySerHisTyrPhe	114		
QY	547	GAACCTGAGTTCAGTAACTAGTCTGCCAGCTGGTGTCTGCTCTCTTCTGCTGAGATATCT	606		
Db	115	GluGluLysTyrSerLysPheCysProAlaGlyCysArgAspIleAlaGlyAspIleSer	134		
QY	607	GGAAACATTCCTCATGGATATAGAGATTCCTGCCCATTTGTCATGGTGTGTGTCATGCA	666		
Db	135	GlyAsnThrLysAspGlyTyrArgAspThrSerLeuLeuCysLysAlaIleHisAla	154		
QY	667	GGAGTAGTCAACACAGCTGGGGCGGCAATCAGTCTTCTAATTAGTAAGGTATTTCCC	726		
Db	155	GlyIleIleThrAspGluLeuGlyGlyHisIleAsnLeuLeuGlnSerLysGlyIleSer	174		
QY	727	TATTATGAAGTCTTTTGGTAAACAGCTCACATCTGTGGTGGGACACTTATCTACAACT	786		
Db	175	HisTyrGluGlyLeuLeuAlaAsnGlyValLeuSerArgHisGlySerLeuSerGluLys	194		
QY	787	CTTTTACATTTAAGACAAGTGGATGTTTGGAACTGCGGATGGAGTCTGGTGTGATC	846		
Db	195	ArgPheLeuPheThrThr-----	200		
QY	847	GCGGATCTCAATAACAGCATCATCTGCTGAGTGGAGTCACTGACCACACAGGGCAAGAG	906		
Db	200	-----	200		
QY	907	AACAGTTGGAACCCCAAAAGCCAGGCTGAAACCAACCTGGACCGCTTGGGCTGCTTT	966		
Db	200	-----	200		
QY	967	GCCACTGATGAATACCAGTGGTTTACAATAGATTTGAATAAGGAAAGAAATAACAGGC	1026		
Db	200	-----	200		
QY	1027	ATTATAACCACTGGATCCACCACATGGTGGAGCAACAATTACTATGTGCTGCTACAGATC	1086		
Db	200	-----	200		
QY	1087	CTGTACAGTGTATGATGGGCGAGAAATGGAGTGTCTACAGAGAGCTGTGTGTGGAGCAAGAT	1146		
Db	201	-----FroglyMet-----	203		
QY	1147	AGATATTTCAAGGAAACAAGATTTATCACAGAGTGTGCGTAAATACTTTTGGCCACCA	1206		
Db	203	-----	203		
QY	1207	ATTATTGCACGTTTTTATTAGAGTGAATCTTACCCTATGGCAGCAGAGAAATGGCATGAAA	1266		
Db	203	-----	203		
QY	1267	ATGGAGCTGCTCGGATGTCAGTTTATTCTTAAAGTGTCTCTCCAAACCTTACTCAACCT	1326		

QY 595 GCTGATATCTGGAACAATTCCTCATGATATAGATTCCTCGCATGTGCTGCT 654
DB 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 655 GGTGTGCATGCAGGAGTAGTGTCAACACAGTTTGGCGGCCAAATCAGT----- 702
DB 182 LysProArgMetGluIleLeuGlnPheLeuThrPheAspLeuGluHisAspProLeu 201
QY 703 -----GTTGTAAATAGTAAAGTAAATCC----- 726
DB 202 GlnValGlyGluGlyAspCysLysTyrAspTyrLeuAspIleThrPheGlyIleProHis 221
QY 727 -----TATTATGAAGTTCITGGCTAACACGTCACATCTGTG 765
DB 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerLysLeuArgSer 241
QY 766 GTGGGACACTTATCTACAAGTCTTTTACATTTAAAGACAAAGT----- 807
DB 242 ThrGlyIleLeuSer-----LeuThrPheHisThrAspMetAlaValAlaLysAsp 258
QY 808 -----GGATGTTAT 816
DB 259 GlyPheSerAlaArgTyrTyrLeuIleHisGlnGluProProGluAsnPheGlnCysAsn 278
QY 817 GGAACACTGGGATGGAGTCTGTGTCATCGCGATCCTCAATAACACGATCATCTGTG 876
DB 279 ValProLeuGlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer--- 297
QY 877 CTGAGTGGACTGACACACAGGCGAAGACAGACAGTTGGAAACCCAAAGACGAGCTG 936
DB 298 -----ThrPheSerAspGlyArgTyrThrProGlnGlnSerArgLeu 311
QY 937 AAAAACTGGACGCGCTTGGCTGCTTTGGCCACTGATGAATACCACTGGTTACAATA 996
DB 312 HisGlyAspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnVal 331
QY 997 GATTGTAATGAAGAAAGAAATAACAGGCAATATACCACTGGA-----TCCACCATG 1050
DB 332 AspleuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGAGCACAAATTAATGTCTGCTGCATACGAATCCTGTACAGTATGATGGCAGAAA 1110
DB 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyLysAsp 371
QY 1111 TGGACTGTGTACAGAGACCTGTGTGTGCGAAGATAGATATTTCAAGAAACAAGAT 1170
DB 372 TrpMetValTyrArgHisGly-----LysAsnHisLysIlePheGlnAlaAsnAsp 389
QY 1171 TATCACCAGNATGCGCTAATAACTTTTGGCCACCAATATTGTCACGTTTATTAGATG 1230
DB 390 AlaThrGluValValLeuAsnLysLeuHisMetProLeuLeuThrArgPheIleArgile 409
QY 1231 AATCCTACCAATGGCAGCAGAAATTTGCCATGAAATGGAGCTGCTCGGATGTCAGTTT 1290
DB 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1291 ATTCCTAAAGTCTGCTCCAAAACCTTACTCAACCTCCACCTCCTCGGAACAGCAATGAC 1350
DB 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCAAAACACTACAGCCCTCCAAAATAGCCAAAGTGTGCGCCCAAAATTT----- 1404
DB 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTrp 460
QY 1405 -----ACGCAACCACTACACCTCGCAGTAGCAATGATTTCTCGCACACAGAA 1455
DB 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgAsnProGln 480
QY 1456 CAACAACACTGCGAGTCTGTATCAGAAATACTACCGTAACCTCCAAATGTAACCAAGAT 1515
DB 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500

QY 1516 GTA-----GGCTGGCTGCAGTTCTTCTCCCTGTGCTG 1548
DB 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPhe 520
QY 1549 GGTATGGTCTCACTACTCTCATTTCTATATTAGTGTGCTGGCAGCTGG----- 1599
DB 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1600 -----AGAAACAGAAAGAAACAACTGAAGCCACC-----TATGACTTTACCTTACTGG 1647
DB 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1648 GACCGGCGAGTGTGGTGAAGGAATGAAGCAGTCTTCTCTGCAAAAGCAGTGGACCAT 1707
DB 559 -----AspIleArgArgPheAspPro----- 565
QY 1708 GAGGAACCCCACTTCGCTATAGCAGCAGGAAGTTAATCACCTGAGTCAACAGAAAGTC 1767
DB 566 -----ValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1768 ACCACAGTGTGCGAGCTGACTCTGCAGAGTAT-----GCTCAGGCACCTGGTAGGAGGA 1821
DB 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1822 ATTGTTGGTACACTTCATCAAGATCTACCTTTAAACCA-----GAAGAAGGA 1869
DB 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAGAAGCAGGCTAT-----GCAGACTAGTATCTTCAACTCA 1908
DB 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerG1 643
QY 1909 CCAGGCGGCAAGTTTATCATGCT-----ATGCTGAACCACTCCCA 1950
DB 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTrpValTyrAspHisAlaLys 663
QY 1951 A---TTACGGGGCTGAGTATGCAACCCCAATCATCATCATGACATGTGAGGCGCACCCCA-- 2005
DB 663 strpLeuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 2006 -----CAACTTCAG----- 2014
DB 676 nAspArgThrPheProAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArgG1 696
QY 2015 -----TTGCTCAGCCCTCCACATCCACTTTCAGGCTACGGGGAACCAA 2058
DB 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 2059 CTTCCCTCCTAGTGGGAACCTTACAATACACTTCTCTC----- 2096
DB 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2097 -----CAGGACTGACAGTGTCTC-----CTCAGCCCGAGCCCGAG 2130
DB 731 lnValValArgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2131 TATGATACCCGGAAGTGGGAACCCAGGCTACTCTCCCGCAGCAGCAATTTGCTTACCAG 2190
DB 751 GlySerGluTrpLysHisGlyArgIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2191 GTGCCACAGACACACAGGAAGTATCAGGACAGGAAGGATGGGATGTGTGTT 2247
DB 770 Ile-----ValPheGluGlyValIleGlyLysArgSerGlyGluIleSerIle 786
RESULT 10
Q8UVRO PRELIMINARY; PRT; 919 AA.
ID Q8UVRO
AC Q8UVRO; 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Neuropilin-2al receptor.
OS Gallus gallus (Chicken).

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:57:21 ; Search time 7.56006 Seconds
(without alignments)
1717.007 Million cell updates/sec

Title: US-10-060-830-1114

Perfect score: 653

Sequence: 1 MPLFLLLLVLLLEDDAGA.....TQVSGAGRGCDVFEKIL 653

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 120991 seqs, 19878514 residues

Word size : 0

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pdb:*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pdb:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pdb:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pdb:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pdb:*

6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pdb:*

7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pdb:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pdb:*

9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pdb:*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pdb:*

11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pdb:*

12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pdb:*

13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pdb:*

14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pdb:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	1.5	677	10	US-09-006-298-21
2	10	1.5	696	10	US-09-925-300-1253
3	9	1.4	150	10	US-09-893-737-230
4	9	1.4	209	10	US-09-893-737-126
5	9	1.4	276	9	US-09-895-913A-356
6	9	1.4	1274	10	US-09-746-491-12
7	8	1.2	86	10	US-09-764-860-544
8	8	1.2	88	10	US-09-731-872-353
9	8	1.2	113	12	US-10-141-938-9
10	8	1.2	113	12	US-10-142-046-9
11	8	1.2	114	8	US-08-927-939-52
12	8	1.2	118	9	US-10-001-876-205
13	8	1.2	140	10	US-09-893-737-118
14	8	1.2	140	10	US-09-746-491-6
15	8	1.2	168	10	US-09-731-872-426
16	8	1.2	184	10	US-09-925-297-737
17	8	1.2	218	10	US-09-746-491-66
18	8	1.2	242	9	US-10-174-590-322
19	8	1.2	242	9	US-10-176-758-322

Sequence 68, Appl
Sequence 322, App
Sequence 984, App
Sequence 155, App
Sequence 22, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 26, Appl
Sequence 12, Appl
Sequence 27, Appl
Sequence 186, App
Sequence 8, Appl
Sequence 336, App
Sequence 336, App
Sequence 4, Appl
Sequence 16, Appl
Sequence 68, Appl
Sequence 4196, Ap
Sequence 2, Appl
Sequence 50, Appl
Sequence 3, Appl
Sequence 20, Appl

20 8 1.2 242 10 US-09-746-491-68
21 8 1.2 242 12 US-10-052-586-322
22 8 1.2 283 10 US-09-925-299-984
23 8 1.2 347 10 US-09-886-055-155
24 8 1.2 396 10 US-09-215-450-22
25 8 1.2 396 10 US-09-953-956-13
26 8 1.2 396 12 US-10-114-464-13
27 8 1.2 398 10 US-09-815-333-2
28 8 1.2 398 10 US-09-842-316-2
29 8 1.2 398 12 US-10-037-616-26
30 8 1.2 400 9 US-09-971-228-12
31 8 1.2 400 12 US-09-812-272-2
32 8 1.2 400 12 US-10-037-616-27
33 8 1.2 471 10 US-09-893-737-186
34 8 1.2 480 9 US-10-084-018-8
35 8 1.2 480 9 US-09-736-457-336
36 8 1.2 480 9 US-09-902-941-336
37 8 1.2 480 9 US-09-849-626-336
38 8 1.2 523 10 US-09-930-218-4
39 8 1.2 527 10 US-09-930-218-16
40 8 1.2 673 10 US-09-782-980-68
41 8 1.2 730 9 US-09-738-626-4196
42 8 1.2 764 9 US-10-045-624B-2
43 8 1.2 829 10 US-09-905-983-50
44 8 1.2 832 10 US-09-824-129-3
45 7 1.1 17 10 US-09-975-901-20

ALIGNMENTS

RESULT 1
US-09-006-298-21 Application US/09006298
Patent No. US20020082224A1
GENERAL INFORMATION:
APPLICANT: Jolly, Douglas J.
APPLICANT: Moore, Margaret D.
APPLICANT: Chada, Sunil
TITLE OF INVENTION: NON-IMMUNOGENIC PRODRUGS AND SELECTABLE
MARKERS FOR USE IN GENE THERAPY
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,298
FILING DATE: 13-JAN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mcmasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 930049.459
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 677 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-006-298-21

Wed Jan 22 14:57:26 2003

us-10-060-830-1114.oli.rapb

Query Match 1.5%; Score 10; DB 10; Length 677;
Best Local Similarity 100.0%; Pred. No. 0.47; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLLLLLL 14
| | | | | | | |
Db 9 LLLLLLLLLL 18

RESULT 2
US-09-925-300-1253
; Sequence 1253, Application US/09925300
; Patent No. US2002011085A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1253
; LENGTH: 696
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; LOCATION: (541)
; NAME/KEY: SITE
; LOCATION: (541)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1253

Query Match 1.5%; Score 10; DB 10; Length 696;
Best Local Similarity 100.0%; Pred. No. 0.49; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLLLLLL 14
| | | | | | | |
Db 28 LLLLLLLLLL 37

RESULT 3
US-09-893-737-230
; Sequence 230, Application US/09893737
; Patent No. US2002011085A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 230
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-230

Query Match 1.4%; Score 9; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LLLLLLLLLL 15
| | | | | | | |
Db 6 LLLLLLLLLL 14

RESULT 4
US-09-893-737-126
; Sequence 126, Application US/09893737
; Patent No. US2002011085A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 126
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-126

Query Match 1.4%; Score 9; DB 10; Length 209;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLLLLLL 13
| | | | | | | |
Db 9 LLLLLLLLLL 17

RESULT 5
US-09-895-913A-356
; Sequence 356, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Comen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 356
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-356

Query Match 1.4%; Score 9; DB 9; Length 276;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLLLLLL 13
| | | | | | | |
Db 16 LLLLLLLLLL 24

RESULT 6
US-09-746-491-12
; Sequence 12, Application US/09746491

```

; Patent No. US20020137202A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-621
; CURRENT APPLICATION NUMBER: US/09/746,491
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: USSN 60/171,329
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1274
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-746-491-12

```

```

Query Match      1.4%; Score 9; DB 10; Length 1274;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches      9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      5 LLLLLLLL 13
        |||||
Db      9 LLLLLLLL 17

```

```

RESULT 7
US-09-764-860-544
; Sequence 544, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 544
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (76)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-860-544

```

```

Query Match      1.2%; Score 8; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 5;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      5 LLLLLLLL 12
        |||||
Db      9 LLLLLLLL 16

```

```

RESULT 8
US-09-731-872-353
; Sequence 353, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78 US3 REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/187,470

```

```

; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 353
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -44...-1
US-09-731-872-353

```

```

Query Match      1.2%; Score 8; DB 10; Length 88;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 PLFLLLLL 9
        |||||
Db     32 PLFLLLLL 39

```

```

RESULT 9
US-10-141-938-9
; Sequence 9, Application US/10141938
; Patent No. US20020132305A1
; GENERAL INFORMATION:
; APPLICANT: Li et al.
; TITLE OF INVENTION: Chemokine Alpha 3
; FILE REFERENCE: PF251D2
; CURRENT APPLICATION NUMBER: US/10/141,938
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 08/816,772
; PRIOR FILING DATE: 1997-03-18
; PRIOR APPLICATION NUMBER: 60/013,615
; PRIOR FILING DATE: 1996-03-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-141-938-9

```

```

Query Match      1.2%; Score 8; DB 12; Length 113;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      8 LLVLLLLL 15
        |||||
Db     21 LLVLLLLL 28

```

```

RESULT 10
US-10-142-046-9
; Sequence 9, Application US/10142046
; Patent No. US20020150994A1
; GENERAL INFORMATION:
; APPLICANT: Li et al.
; TITLE OF INVENTION: Chemokine Alpha 3
; FILE REFERENCE: PF251D3
; CURRENT APPLICATION NUMBER: US/10/142,046
; CURRENT FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 08/816,772
; PRIOR FILING DATE: 1997-03-18
; PRIOR APPLICATION NUMBER: 60/013,615
; PRIOR FILING DATE: 1996-03-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-142-046-9

```

Query Match 1.2%; Score 8; DB 12; Length 113;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
| | | | | | | |
Db 21 LLVLLLLL 28

RESULT 11
US-08-927-939-52
; Sequence 52, Application US/08927939
; Patent No. US20010006640A1
; GENERAL INFORMATION:
; APPLICANT: Grainger, David J.
; APPLICANT: Tatalick, Lauen Marie
; TITLE OF INVENTION: Compounds and methods to inhibit or
; FILE REFERENCE: 295.022US1
; CURRENT APPLICATION NUMBER: US/08/927,939
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 52
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-927-939-52

Query Match 1.2%; Score 8; DB 8; Length 114;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
| | | | | | | |
Db 21 LLVLLLLL 28

RESULT 12
US-10-001-876-205
; Sequence 205, Application US/10001876
; Patent No. US2002017140A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caferkey, Robert
; APPLICANT: Ali, Shujath
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and
; FILE REFERENCE: DEX-0285
; CURRENT APPLICATION NUMBER: US/10/001,876
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,186
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 205
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Homo saplen
US-10-001-876-205

Query Match 1.2%; Score 8; DB 9; Length 118;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
| | | | | | | |
Db 27 LLVLLLLL 34

RESULT 13
US-09-893-737-118
; Sequence 118, Application US/09893737
; Patent No. US20020110855A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 118
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-118

Query Match 1.2%; Score 8; DB 10; Length 140;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
| | | | | | | |
Db 4 LLVLLLLL 11

RESULT 14
US-09-746-491-6
; Sequence 6, Application US/09746491
; Patent No. US20020137202A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-621
; CURRENT APPLICATION NUMBER: US/09/746,491
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: USSN 60/171,329
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-746-491-6

Query Match 1.2%; Score 8; DB 10; Length 140;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
| | | | | | | |
Db 4 LLVLLLLL 11

RESULT 15
US-09-731-872-426
; Sequence 426, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78 US3 REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,629

```
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 426
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-731-872-426

Query Match      1.2%; Score 8; DB 10; Length 168;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 LLVLLLLL 15
        |||||
Db      151 LLVLLLLL 158

Search completed: January 21, 2003, 10:00:39
Job time : 8.56006 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 21, 2003, 10:00:41 ; Search time 253.89 Seconds
(without alignments)
6877.663 Million cell updates/sec
Title: US-10-060-830-1115
Perfect score: 60
Sequence: 1 ctgtctctcggagcgtgg.....gtggacacactgtactagc 60
Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 2054640 seqs, 14551402878 residues
Word size : 0
Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	60	100.0	1907	6	AX405795	AX405795 Sequence
2	60	100.0	2020	9	AK056350	AK056350 Homo sapi
3	60	100.0	2939	9	AF387547	AF387547 Homo sapi
4	60	100.0	3858	9	BC029658	BC029658 Homo sapi
5	60	100.0	6093	9	AB073146	AB073146 Homo sapi
6	37	61.7	2642	10	AF387548	AF387548 Mus muscu
7	35	58.3	675	9	HS332516	AJ332516 Homo sapi
8	35	58.3	35018	9	AC026303	AC026303 Homo sapi
9	35	58.3	146515	2	AC025661	AC025661 Homo sapi
10	35	58.3	164216	2	AC091213	AC091213 Homo sapi
11	32	53.3	586	9	BC007117	BC007117 Homo sapi
12	32	53.3	2339	10	AF387549	AF387549 Rattus no
13	32	53.3	61344	2	AC073443	AC073443 Homo sapi
14	32	53.3	164216	2	AC091213	AC091213 Homo sapi
15	32	53.3	192247	2	AC013497	AC013497 Homo sapi
16	32	53.3	193623	9	AC106728	AC106728 Homo sapi
17	32	53.3	203102	2	AC126975	AC126975 Rattus no
18	29	48.3	1388	9	HUMCUB1	D29810 Human mRNA
19	28	46.7	59871	2	AC101046	AC101046 Mus muscu
20	20	33.3	76528	2	AC102949	AC102949 Homo sapi
21	20	33.3	76528	2	AC102949	AC102949 Homo sapi
22	20	33.3	175910	2	AC091004	AC091004 Homo sapi
23	20	33.3	202051	2	AC060775	AC060775 Homo sapi
24	19	31.7	100324	2	AC112025	AC112025 Rattus no
25	19	31.7	174223	2	AC126662	AC126662 Rattus no
26	18	30.0	16178	9	HS8569	AJ006945 Human P2Y
27	18	30.0	73557	2	AC093631	AC093631 Homo sapi
28	18	30.0	103718	9	AL139175	AL139175 Human DNA
29	18	30.0	122322	9	AC002448	AC002448 Homo sapi
30	18	30.0	134760	9	AC099484	AC099484 Homo sapi
31	18	30.0	146059	2	AC019030	AC019030 Homo sapi
32	18	30.0	146283	2	AC022277	AC022277 Homo sapi
33	18	30.0	146647	2	AC128191	AC128191 Rattus no
34	18	30.0	147575	2	AC117139	AC117139 Rattus no
35	18	30.0	147856	2	AC079116	AC079116 Homo sapi
36	18	30.0	153402	2	AC069485	AC069485 Homo sapi
37	18	30.0	153667	9	AC109439	AC109439 Homo sapi
38	18	30.0	157452	2	AC010190	AC010190 Homo sapi
39	18	30.0	161679	9	AL592435	AL592435 Human DNA
40	18	30.0	167611	9	AL583859	AL583859 Human DNA
41	18	30.0	176170	2	AC021662	AC021662 Homo sapi
42	18	30.0	176251	9	AC013251	AC013251 Homo sapi
43	18	30.0	178015	2	AC108113	AC108113 Homo sapi
44	18	30.0	180064	9	AC025265	AC025265 Homo sapi
45	18	30.0	180525	2	AC063946	AC063946 Homo sapi

ALIGNMENTS

RESULT 1
AX405795
LOCUS AX405795
DEFINITION Sequence 210 from Patent WO0222660.
ACCESSION AX405795
VERSION AX405795.1 GI:21439029
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,
Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.
TITLE Novel nucleic acids and polypeptides

RP. SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; Z81512; CAB04172.1; --
 DR InterPro; IPR000524; HTH_GntR.
 DR PROSITE; PS00043; HTH_GNTR_FAMILY; UNKNOWN_1.
 SQ SEQUENCE 3147 AA; 354192 MW; 1F411CA0A4BFE6A CRC64;

Alignment Scores:
 Pred. No.: 1.34 Length: 3147
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.09% Indels: 0
 DB: 5 Gaps: 0

US-10-060-830-4 (1-300) x Q9XV66 (1-3147)

Oy 145 TCACCTGAGTCCAGAGAGTACCAC 171
 AC Q9J842; PRT; 85 AA.
 Db 2403 SerProGluSerLysArgSerHis 2411
 |||||

RESULT 7

Q9J842
 ID Q9J842 PRELIMINARY; PRT; 85 AA.
 AC Q9J842;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ORF95.
 OS Spodoptera exigua nucleopolyhedrovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 OC Nucleopolyhedrovirus.
 OX NCBI_TaxID=10454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20036646; PubMed=10567663;
 RA IJkel W.F., van Strien E.A., Heldens J.G., Broer R., Zuidema D.,
 RA Goldbach R.W., Vlask J.M.;
 RT "Sequence and organization of the spodoptera exigua multicapsid
 RT nucleopolyhedrovirus genome.";
 RL J. Gen. Virol. 80:3289-3304(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA IJkel W.F.J., van Strien E.A., Heldens J.G.M., Broer R., Zuidema D.,
 RA Goldbach R.W., Vlask J.M.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF169823; AAF33624.1; --
 SQ SEQUENCE 85 AA; 9972 MW; DFDIC18B9B79365A CRC64;

Alignment Scores:
 Pred. No.: 23 Length: 85
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 12 Gaps: 0

US-10-060-830-4 (1-300) x Q9J842 (1-85)

Oy 278 CTGCTCTTTCTCTCTCTCTCTGTT 255
 |||||

Db 26 LeuLeuPheLeuLeuVal 33
 |||||

RESULT 8

Q9GRU7
 ID Q9GRU7 PRELIMINARY; PRT; 103 AA.
 AC Q9GRU7;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE Hypothetical secreted protein precursor.
 GN SPS20.
 OS Heterodera schachtlii.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 OC Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
 OX NCBI_TaxID=97005;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA De Meutter J.J.;
 RT "Cloning of a gene from Heterodera schachtlii coding for a secreted
 RT protein.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ297919; CAC15549.1; --
 KW Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 103 POTENTIAL.
 SQ SEQUENCE 103 AA; 11834 MW; 7B5888C2A7BE06A6 CRC64;

Alignment Scores:

Pred. No.: 22.4 Length: 103
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 5 Gaps: 0

US-10-060-830-4 (1-300) x Q9GRU7 (1-103)

Oy 284 CATAGCGTCTCTCTCTCTCTT 261
 |||||

Db 7 HisSerLeuLeuPheLeuLeu 14
 |||||

RESULT 9

Q9UWS7
 ID Q9UWS7 PRELIMINARY; PRT; 104 AA.
 AC Q9UWS7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein SSO0521.
 GN SSO0521 ORF-C22.030.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aways M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 DR EMBL; Y18930; CAB57778.1; --
 DR EMBL; AE006682; AAK40839.1; --
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 104 AA; 11535 MW; 1C0E2E3D28DF66B6 CRC64;

Alignment Scores:

Pred. No.: 22.4 Length: 104
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.08% Indels: 0
 DB: 17 Gaps: 0

US-10-060-830-4 (1-300) x Q9UWS7 (1-104)

```

QY 206 CCACGTGTAGGAGGATTTGGT 229
Db 86 ProLeuValGlyIleValcIy 93

RESULT 10
Q9DCH7 PRELIMINARY; PRT; 118 AA.
ID Q9DCH7
AC Q9DCH7;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE 0610037K01Rik protein.
GN 0610037K01Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Kaschiwa H.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK008151; BAB25496.1;
DR MGD; MGI:1917364; 2010007L08Rik.
SQ SEQUENCE 137 AA; 15676 MW; E0341C56E87C916E CRC64;

Alignment Scores:
Pred. No.: 21.7 Length: 137
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 11 Gaps: 0

US-10-060-830-4 (1-300) x Q9D8C7 (1-137)
QY 85 AAGAACTGCTTCATTCCTTCCCA 62
Db 12 LysLysLeuLeuHisSerPhePro 19

RESULT 12
Q976P5 PRELIMINARY; PRT; 231 AA.
ID Q976P5
AC Q976P5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical protein ST0146.
GN ST0146.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudo H., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermophilic
RT Crenarchaeon, Sulfolobus tokodaii strain 7."
RL DNA Res. 8:123-140(2001).
DR EMBL; AF000981; BAB65101.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 231 AA; 25529 MW; 16AB11D74C2A190C CRC64;

Alignment Scores:

```

Db 1403 ACTTCGGGGGAGATGATGTT 1423

RESULT 10
US-09-864-761-12562
; Sequence 12562, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12562
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
; US-09-864-761-12562

Alignment Scores:
Pred. No.: 8,33e-32 Length: 568
Score: 428.00 Matches: 78
Percent Similarity: 97.59% Conservative: 3
Best Local Similarity: 93.98% Mismatches: 0

Db 1403 ACTTCGGGGGAGATGATGTT 1423

712 ----- 712
Qy GlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProProArg 421
Db 712 ----- 712
Qy AsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaPro 441
Db 712 ----- 712
Qy LysPheThrClnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGln 461
Db 712 ----- 712
Qy ThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAspVal 481
Db 713 ----- 712
Qy AlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuLeuLeu 501
Db 728 GCGATTCCATCAGTGATC-----TTTCATCGCCCTCCTCTGACTGGAATGGGATCTTT 781
Qy LeuValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrAspLeu 521
Db 782 GCAATCTGT-----AGAAAGAGGAAAGAAAGAAAT----- 814
Qy TrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeu 537
Db 815 CCATATGTGTGACGTGACGCTCAGAAACAGCGTGTGGAGACAGATTAAATATCCCTTT 874
Qy ProAlaLysAlaValAspHisGluGluThrProValArgTyrSerSerGlu----- 555
Db 875 -----GCCAGGCATCAGTCGACGCGAATTACC 901
Qy ValAsnHisLeuSerProArgGluValThr-----ThrValLeuGlnAlaAspSer 572
Db 902 ATCAGCTATGACATGAAAGAGATGACACAAAGTTGGATCTCATCATGATGATG 961
Qy AlaGluThrAlaGlnProLeuValGlyGlyLeuValGlyThrLeuHisGlnArgSerThr 592
Db 962 GCAGATTATCAGCAGCTCTCATGATGGCAGGCACATCGCGAGAAAGGCTCTACC 1021
Qy PheLysPro-----GluGlu-----GlyLysGluAlaGlyTyr 603
Db 1022 TTCGAGCCCATGGACACAGACACTGAGGAGGTGACAGTGAACACTGAGCGCGCCAC 1081
Qy AlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluPro 623
Db 1082 -----TATGACTGTCTTCACCGCCGCGCGCCGCGCATGATGACGCACTGCCT 1126
Qy LeuProIleThrGlyProAlaThrProIleThr----- 636
Db 1127 TTGAGGCACTCAGAACCTGATGATGACCACTATCTGAGCGGCGACCTGCTGCGAGCT 1186
Qy 637 -----MetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSer 653
Db 1187 CACACCTTCTCCACACAGCGGTACTCGAGTCCCTCGG-----CCAGGCCCACTCACAAA 1243
Qy ThrPheLysAlaThrGlyAsnGlnProProProLeuValGlyThrThrAsnThrLeuLeu 673
Db 1244 CACTCCCATCTCTGAGGCGTTTCCCTGCTGCTACAGGAGCCACCCAGGTGGAAGCTAT 1303
Qy SerArgThrAspSerCysSerSerAlaGlnAlaGlnThrAspThrProLysAlaGlyLys 693
Db 1304 CAGAGGCCACGACCCCAAGCCCTGTGGTGGTGGTATGACAAAGCCT---GCTGCTAGC 1360
Qy ProGlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluVal 713
Db 1361 AGCTTCTTGGACACGACAGAC-----CCAGCCTCTCAGTCACAGATG 1402
Qy 714 SerGlyAlaGlyArgAspGly 720

Query Match: 11.01% Indels: 2
DB: 10 Gaps: 1

US-10-060-830-3 (1-729) x US-09-864-761-12562 (1-568)

Qy 22 GlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerile 41
Db 230 AAGGTATGATGATGGACACACTGCTACTAGCCCTGAGAGTGGACCCCTACATCCATA 289
Qy 42 AsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluLeuArgValLysMet 61
Db 230 AACTACCACACAGACTATCCCAACAGCACTCTTGTGAATGGAGATCCGCTGAAGATG 349
Qy 62 GlyGluArgValArgIleLysPheGlyAspPheAspPheGlyLeuAspSerAspSerCysHis 81
Db 350 GGAGAGAGAGTTCGATCAAAATTTGGTGACTTGACATTTGAAGATTCTGATCTGTGCAC 409
Qy 82 PheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys--- 100
Db 410 TTTAATTACTTGAGAAATTTAATGGAATTCGAGTCCAGCAACTGAAATAGGTAGGACT 469
Qy 101 ---TyrCys 102
Db 470 TTTTCTTCT 478

RESULT 11

US-10-003-132-13
; Sequence 13, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate nucleotide sequence
; NAME/KEY: misc.feature
; LOCATION: (1)...(1509)
; OTHER INFORMATION: n = A,T,C or G
US-10-003-132-13

Alignment Scores:
Pred. No.: 5,37e-30 Length: 1509
Score: 415.50 Matches: 100
Percent Similarity: 43.89% Conservative: 40
Best Local Similarity: 31.35% Mismatches: 130
Query Match: 10.69% Indels: 49
DB: 9 Gaps: 10

US-10-060-830-3 (1-729) x US-10-003-132-13 (1-1509)

Qy 21 GlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSer 40
Db 79 GARYTNGGAYGGTGGNCGAYATHTGNACNWSNARGAYWSNGNACNATGACNWSN 138
Qy 41 IleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluIleArgValLys 60
Db 139 ARAAYATCCNGNACNATYACCNAAATAYACNAGTNTGAGAAATATHTACNCTNCCN 198
Qy 61 MetGlyGluArgValArgIleLysPheGlyAspPheAspPheGlyLeuAspSerAspSerCys 80
Db 199 AARGGNAARMGNTNATHTYTNMGNTNGNGAYTNAAYATHGAR---WSNAARACNTGY 255

Qy 81 HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys 100
Db 256 GCNWSNGAY-----TAYTNTNTYWSNWSNCGNACNCGAYCATATGGNCCN 303
Qy 101 TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr 120
Db 304 TAYTGYGSGN--WSNTGGCNGTNCNCAARGARVNTMGNTNAAAYWSNAAYGARGTNACN 360
Qy 121 LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal 140
Db 361 GTNTYNTTYAARWNGGNGWSNCAATATWSNNGNMGNGNTTYTNTNACNATAYGCMWSN 420
Qy 141 IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu 160
Db 421 WNGAYCAYCCNGAYTNTACNTGYTTNGARMGNGNWSNCAITATTTTGARGAAR 480
Qy 161 PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIle 180
Db 481 TAYWSNAAATYTCYCCNGCNGNTGYMGNGAYATHCNGCGNGAYATHSNGGNAAYACN 540
Qy 181 ProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValVal 200
Db 541 AARGAYGGTAYMGNGAYACNWSNTYNTNTGYAARGCNGCNATHCAYCGCNGNATHATH 600
Qy 201 SerAsnThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGlu 220
Db 601 ACNGAYGARYTNGGNGNCAATHAAYTNTTNCARNSNAARGNATHWSNCAITAYGAR 660
Qy 221 SerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr 240
Db 661 GGNTYNTNGCNAAYGGNGTNTYWSNMGNCAYCGNWSNYTWSNGAARMGNTTYTN 720
Qy 241 PheLysThrSerGly-----Cys 246
Db 721 TTYACNACCCNGGNGATGAAYATACNACNGTNGCNATHCCNWSNNGTNTATHTTATHGCGN 780
Qy 247 TyrGlyThrLeuGlyMetGluSerGlyValIleAla----- 258
Db 781 YTNTYNTNACNGGNGATG-----GGNATHTYGCGNATHTYGMGNAARMGNAARMARAR 834
Qy 259 ---AspProGlnIleThrAlaSerSerValLeuGluTrpThrAspHisThrGlyGlnGlu 277
Db 835 GGNAAYCCNTAYGTNWSNCGNGAYCGN-----CARAARACNCGN----- 873
Qy 278 AsnSerTrpLysProLysLysAlaArgLeuLysLysProGlyProTrpAlaAlaPhe 297
Db 874 ---TGYTGAAR-----CARATHAARTAYCCN-----TTT 900
Qy 298 AlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThr 316
Db 901 GCNMGNCAYCARWSNACNGARTTYACNATHWSNTAYGAYAGARARARATGACN 957

RESULT 12

US-09-864-761-29163
; Sequence 29163, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Weosheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; FILE REFERENCE: Aescmca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03

```
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 4, 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29163
; LENGTH: 228
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
; OTHER INFORMATION: NT HIT: D29810.1, EVALUATE 1.00e-125
; OTHER INFORMATION: EST_HUMAN HIT: A1565996.1, EVALUATE 1.00e-125
; OTHER INFORMATION: SWISSPROT HIT: P98063, EVALUATE 1.00e-09
US-09-864-761-29163

Alignment Scores:
Pred. NO.: 6.99e-31 Length: 228
Score: 412.00 Matches: 75
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.60% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-3 (1-729) x US-09-864-761-29163 (1-228)
Qy 24 AspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerIleAsnTyr 43
Db 3 GATGATGTGACACACTGTACTAGGCCCTGAGAGTGAACCCCTTACATCCATAAACTAC 62
Qy 44 ProGlnThrTyrProAsnSerThrValCysGluTyrGluLeuArgValLysMetGlyGlu 63
Db 63 CCACAGACCTATCCACACAGCAGCTGTTGTCATYGGGAGATCCGTTAAAGATGGGAGAG 122
Qy 64 ArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCysHisPheAsn 83
Db 123 AGAGTTCGATCAAAATTTGGTGACTTTCACATTTGAAGATTCGTATCTTGTCACATTAAT 182
Qy 84 TyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIle 98
Db 84 TTTCTCTGCGCTACGAGATCTTCAAGACAGGCTCT-----GAAATGCTCA 546

Db 183 TACTTGAGATTTATATGGAATTGGAGTGGAGTACGAGCACTGAATA 227

RESULT 13
US-10-104-440-3
; Sequence 3, Application US/10104440
; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Quan
; TITLE OF INVENTION: ANTAGONISTS OF NEUROPILIN RECEPTOR FUNCTION AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 48802 C
; CURRENT APPLICATION NUMBER: US/10/104,440
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/580,803
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/069,155
; PRIOR FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 60/069,687
; PRIOR FILING DATE: 1997-12-29
; PRIOR APPLICATION NUMBER: 60/078,541
; PRIOR FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 3404
; TYPE: DNA
; ORGANISM: human
US-10-104-440-3

Alignment Scores:
Pred. NO.: 4.76e-29 Length: 3404
Score: 411.00 Matches: 191
Percent Similarity: 36.56% Conservative: 104
Best Local Similarity: 23.67% Mismatches: 342
Query Match: 10.57% Indels: 170
DB: 12 Gaps: 31

US-10-060-830-3 (1-729) x US-10-104-440-3 (1-3404)
Qy 1 MetProLeuPheLeuLeu-----LeuLeuLeuValLeuLeuLeuLeuGluAspAla 18
Db 97 ATGATATGTTTCTCTCACCTGGGTTTCTTAGCCCTCTACTTTTCAAGACACCAAGTG 156
Qy 19 GlyAlaGlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeu 38
Db 157 AGAGGCCAACACAGACCCACCGTGGCGAGTCTGTTGAATTCGAAAGATCTGCTGCTATATC 216
Qy 39 ThrSerIleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTyrGluIleArg 58
Db 217 ACCTCTCCCGGTATACCCCGAGCTACCCCTCCACACAGAACTGCGAGTGGATGTTTAC 276
Qy 59 Val---LysMetGlyGluArgValArgIleLysPheGly---AspPheAspIleGluAsp 76
Db 277 GCGCCCGAACCCACACAGAGATTCCTCAACTCAACCCCTCACTTTGAATTCGAGAG 336
Qy 77 SerAspSerCysHisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThr 96
Db 337 CACGAC---TGCAGTATGACTTTATCGAGATTCGGGATGGGACAGTGAATCCGACAG 393
Qy 97 GluIleGlyLysTyrCysGlyLeuGlyLeuGlnMetAsn-----HisSerIleGlu 113
Db 394 CTCCTGGGCAAAACACTGTGG-----AAATCGCCCGCCGCCACCATCATC 438
Qy 114 SerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGly 133
Db 439 TCCTCGGCTCCATCTCTACATCAAGTTCACCTCCCTACTCGCCGCGGCGGCGAGC 498
Qy 134 PheLeuAlaSerTyrSerValIleAspLysGlnAspLeuLeuIleThrCysLeuAspThrAla 153
Db 499 TTCTCTGCGCTACGAGATCTTCAAGACAGGCTCT-----GAAATGCTCA 546
```

QY 154 SerAsnProGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPhe 173
Db 547 AAAAATTCACAAACCCAGGAGGACCAATGCTCTCGG-----TTT 591
QY 174 AlaGluLeuSerGlyThrIleProHisGlyTyrArgAspSerProLeuCysMetAla 193
Db 592 CCTGAG-----AAGTATCCACACAACCTTG--GACTGCACCTTTACCATCCTGCC 639
QY 194 GlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSer----- 209
Db 640 AAACCCAGATGGAGATCATCTCGAGTCTCTGATCTTTCAGCTGGAGCATGACCTTTG 699
QY 210 -----ValValIleSerLysLysIlePro--- 217
Db 700 CAGGTGGAGAGGGGACTGCAAGTACAGTGGCTGGACATCTGGGATGGCATTCACAT 759
QY 218 -----TyrTyrGluSerSerLeuAlaAsnValThrSerVal 230
Db 760 GTTGCCCTGATGGCAAGTACTGTGGGACCAAAACACCTCTGAATTCGTTTCATCG 819
QY 231 ValGlyHisLeuSerThrSerLeuPheThr----- 240
Db 820 ACGGGATCTCTCCTGACCTTTCACACGACATGCGGTGGCCAAAGGATGGCTTCTCT 879
QY 241 -----PheLysThrSerGlyCysTyrGlyThrLeu 250
Db 880 GCGGTACTACCTGGTCCACAGACCACTAGACAACTTTCAGTGCATGTTCCTCTG 939
QY 251 GlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerValLeuGluTrp 270
Db 940 GCGATGGATCTGCGCGGATGCTTAATGAACAGATCAGTGCCTCATCT----- 987
QY 271 ThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLysLysPro 290
Db 988 -----ACCTACTGTGAGGGAGTGGACCTCTCAACAAAGCCGCTCATGGTGTAT 1038
QY 291 GlyProProAlaAlaPheAlaThrAspGluTyrGlnIleLeuGlnIleAspLeuAsn 310
Db 1039 GACAATGGCTGGACCCCACTGGATTCACCAAGGAGTATCTCCAGGTGGACCTGGC 1098
QY 311 LysGluLysLysIleThrGlyIleThrGly-----SerThrMetValGluHis 328
Db 1099 TTTTAAACCATGCTCAGCGCATCCACACAGAGGAGGATTTCCAGGGAACACAGAA 1158
QY 329 AsnTyrTrpValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTrpThrVal 348
Db 1159 GCTACTACGTAATCTCAAGCTGGAAGTCACTAATGAGGAGGACTGGATGTG 1218
QY 349 TyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyrHisGln 368
Db 1219 TACCGGCATGGC-----AAAAACCAAGGTATTTCAAGCAACACAGATGCAACTGAG 1272
QY 369 AspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgValAsnProThr 388
Db 1273 CTGGTCTTGAACAACCTCCACGCTCCATGCTGACAGAGTTGTAGAAATCCGCTCAG 1332
QY 389 GlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheIleProLys 408
Db 1333 ACCTGGCACTAGGTATCGCCCTCGGCTGGAGCTCTTCGGCTGC----- 1377
QY 409 GlyArgProProLysLeuThrGlnProProProArgAsnSerAsnAspLeuLysAsn 428
Db 1378 -----CGGTGCACAGATGCTCCCTGCTCCACATGCTGGGGATGCTCTCAGGC 1425
QY 429 ThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPhe----- 443
Db 1426 CTCATTGCAGACTCCAGATCTCCGCTCTTCCACCCAGGAATACCTCTGGAGCCCACT 1485
QY 444 ThrGlnProGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThr 463
Db 1486 GCAGCCGCTGTCAGACGCGCTGGGTGGTTCCTCGAATCCCTCAGCCCGACGCC 1545
QY 464 AlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAspVal----- 481

Db 1546 GGTGAGGAGTGGCTTTCAGGTAGATCTGGGAACACCCAGACAGTGAAGGTGTCTATCATC 1605
QY 482 -----AlaLeuAlaValLeuValProValLeuValMetVal 494
Db 1606 CAGGAGCCCGGAGGAGACAGTACTCTCTGGAAGCCAGAGCATTTGTCCGCAAG 1665
QY 495 LeuThrThrLeuIleLeuValCysAlaThrHisTrp-----Arg 509
Db 1666 TTCAAACTCTCCACAGCTAAACGGCAAGACTGGGAATATCATTCAGGACCCAGGACC 1725
QY 510 AsnArgLysLysLysThrGluGlyThr-----TyrAspLeuProTyrTrpAspArgAla 527
Db 1726 CAGCAGCAAAAGCTGTTTGAAGGGAACATGCACTATCACACCCCT----- 1770
QY 528 GlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThr 547
Db 1771 -----GACATCCGAAGGTTTACCCCATTCGCGCACAGTATGTCGGGTATAC 1818
QY 548 ProValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrThrVal 567
Db 1819 CCGGAGAGGTGGTCCCGCGGGGATTTGGGATCGG-----CTGGAGGTG 1863
QY 568 LeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeu 587
Db 1864 CTGGCTGTGACTGCACAGAC---TCCAAGCCACGGTAGAGACGCTGGGACCCACTGTG 1920
QY 588 HisGlnArgSerThrPheLysPro-----GluGluGlyLysGluAlaGly--- 602
Db 1921 AAGAGCGAAGAGACAACACCCCTTACCCACCAAGAGAGGAGGAGAGTGTGGGAG 1980
QY 603 -----TyrAlaAspLeuAspProTyrAsnSerProGly-GlnGluValTyrHisAl 619
Db 1981 AACTGCAGCTTGAGGATGACAAAGATTTGCAGCTCCCTCGGATTCATTCGCAACTTC 2040
QY 619 aTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetAspMe 639
Db 2041 GATTTCTCGAGGAGCCCTGTGGTTGA-----TGTATGACCAT 2079
QY 639 t-----SerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLy 656
Db 2080 GCCAAGTGGCTCCGAGCACCTGGGCCAGCAGCTCCAGCCCAACGACCGAGCTTCCA 2139
QY 656 salatThrGlyAsnGlnProProProLeuValGlyThrTyrAsnThrLeuLeuSerArgTh 676
Db 2140 GATCAGAGA----- 2149
QY 676 rAspSerCysSer-----SerAlaGlnAlaGlnTyrAspThrProLysAlaGly-L 693
Db 2150 -ATTTCTTGGCTGTCAGAGTGACAGCCAGAGAGAGGCGCATGATGCCCGCTCATCAGC 2208
QY 693 ysProGlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluV 713
Db 2209 CCCCTGTCCACCTGCCCGGAGCCCGGTGTGATGGAGTTCCA-----G 2253
QY 713 alSerGlyAlaGlyArg 718
Db 2254 TACCAGGCCACCGGGCGG 2270
RESULT 14
US-10-104-440-1
; Sequence 1, Application US/1010440
; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Quan
; TITLE OF INVENTION: ANTAGONISTS OF NEUROPILIN RECEPTOR FUNCTION AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 48802 C
; CURRENT APPLICATION NUMBER: US/10/104,440
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/580,803

100

[illegible]

AC Q33800;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytochrome B, COB OR CYTB.
 GN MTCYB OR COB OR CYTB.
 OS Antechinus naso (Antechinus habbema).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.
 OX NCBI_TaxID=38774;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Armstrong L.A., Krajewski C., Westerman M.;
 RT "Phylogeny of the dasyurid marsupial genus Antechinus based on
 RT cytochrome b, 12S rRNA, and protamine P1 genes.";
 RL J. Mammal. 79:1379-1389(1998).
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS.
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN.
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U23461; AAB91425.1; -
 DR InterPro; IPR000179; Cyt.b_b6.
 DR Pfam; PF00032; cytochrome_b_c_1.
 DR Pfam; PF00033; cytochrome_b_n_1.
 DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE; PS00193; CYTOCHROME_B_OO; 1.
 KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
 KW Heme.
 FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 182 182 IRON 2 (HEME B562 AXIAL LIGAND).
 FT METAL 196 196 IRON 1 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 381 AA; 42899 MW; 61C44C2BF7921E90 CRC64;
 Query Match 1.4%; Score 9; DB 1; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 6 LLLVLVLLL 14
 | | | | |
 Db 234 LLLVLVLLL 242
 RESULT 8
 CYB_DASHA STANDARD; PRT; 381 AA.
 AC Q34321;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Cytochrome B.
 GN MTCYB OR COB OR CYTB.
 OS Dasyurus hallucatus (Satanellus/northern quoll).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Dasyurus.
 OX NCBI_TaxID=9280;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=93096825; PubMed=1361058;
 RA Krajewski C., Driskell A.C., Baverstock P.R., Braun M.J.;
 RT "Phylogenetic relationships of the thylacine (Mammalia: Thylacinidae)
 RT among dasyuroid marsupials: evidence from cytochrome b DNA
 RT sequences";
 RL Proc. R. Soc. Lond., B, Biol. Sci. 250:19-27(1992).
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS.
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN.
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M99460; AAB40405.1; -
 DR InterPro; IPR000179; Cyt.b_b6.
 DR Pfam; PF00032; cytochrome_b_c_1.
 DR Pfam; PF00033; cytochrome_b_n_1.
 DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE; PS00193; CYTOCHROME_B_OO; 1.
 KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
 KW Heme.
 FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 182 182 IRON 2 (HEME B562 AXIAL LIGAND).
 FT METAL 196 196 IRON 1 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 381 AA; 42914 MW; 49320F01FFABFB00 CRC64;
 Query Match 1.4%; Score 9; DB 1; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 6 LLLVLVLLL 14
 | | | | |
 Db 234 LLLVLVLLL 242
 RESULT 9
 CYB_NEOLO STANDARD; PRT; 381 AA.
 AC Q35157;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Cytochrome B.
 GN MTCYB OR COB OR CYTB.
 OS Neophascogale lorentzii (Long-clawed marsupial mouse).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Neophascogale.
 OX NCBI_TaxID=32551;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Krajewski C., Painter J., Buckley L., Westerman M.;
 RT "Phylogenetic structure of the marsupial family Dasyuridae based on
 RT cytochrome-b DNA sequences";
 RL J. Mammal. Evol. 2:25-35(1994).
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS.
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY

CC BOUND TO THE PROTEIN.
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U07585; AAB88761.1; -
 CC InterPro: IPR000179; Cyt_b_b6.
 CC Pfam: PF00032; Cytochrome_b_C; 1.
 CC Pfam: PF00033; Cytochrome_b_N; 1.
 CC PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 CC PROSITE: PS00193; CYTOCHROME_B_QO; 1.
 CC Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
 CC KW Heme.
 CC FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
 CC FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
 CC FT METAL 182 182 IRON 2 (HEME B562 AXIAL LIGAND).
 CC FT METAL 196 196 IRON 1 (HEME B566 AXIAL LIGAND).
 CC SQ SEQUENCE 381 AA; 42784 MW; 121C23A0B0B45A7E CRC64;
 CC
 CC Query Match 1.4%; Score 9; DB 1; Length 381;
 CC Best Local Similarity 100.0%; Pred. No. 1.1;
 CC Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 6 LLLLVLLLL 14
 CC | | | | | | | |
 CC DB 234 LLLLVLLLL 242
 CC
 CC RESULT 10
 CC CYB_PHATA
 CC ID CYB_PHATA STANDARD; PRT; 381 AA.
 CC AC Q35673;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
 CC DE Cytochrome B.
 CC GN MTCYB OR COB OR CYTB.
 CC OS Phascolale tapoatafa (Common wambenger).
 CC OG Mitochondrion.
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Phascogale.
 CC OX NCBI_TaxID=9293;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=93096825; PubMed=1361058;
 CC RA Krajewski C., Driskell A.C., Baverstock P.R., Braun M.J.;
 CC RT "Phylogenetic relationships of the thylacine (Mammalia: Thylacinidae)
 CC RT among dasyuroid marsupials: evidence from cytochrome b DNA
 CC RT sequences.";
 CC RL Proc. R. Soc. Lond., B, Biol. Sci. 250:19-27(1992).
 CC CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC CC COUPLED TO ATP SYNTHESIS.
 CC CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC CC BOUND TO THE PROTEIN.
 CC CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
 CC CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 CC CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M99459; AAB40404.1; -
 CC InterPro: IPR000179; Cyt_b_b6.
 CC Pfam: PF00032; Cytochrome_b_C; 1.
 CC Pfam: PF00033; Cytochrome_b_N; 1.
 CC PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
 CC PROSITE: PS00193; CYTOCHROME_B_QO; 1.
 CC Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
 CC KW Heme.
 CC FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
 CC FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
 CC FT METAL 182 182 IRON 2 (HEME B562 AXIAL LIGAND).
 CC FT METAL 196 196 IRON 1 (HEME B566 AXIAL LIGAND).
 CC SQ SEQUENCE 381 AA; 42755 MW; B8F15BEF4D5B41C7 CRC64;
 CC
 CC Query Match 1.4%; Score 9; DB 1; Length 381;
 CC Best Local Similarity 100.0%; Pred. No. 1.1;
 CC Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 6 LLLLVLLLL 14
 CC | | | | | | | |
 CC DB 234 LLLLVLLLL 242
 CC
 CC RESULT 11
 CC CATE_CAVPO
 CC ID CATE_CAVPO STANDARD; PRT; 391 AA.
 CC AC P25796;
 CC DT 01-MAY-1992 (Rel. 22, Created)
 CC DT 01-MAY-1992 (Rel. 22, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE Cathepsin E precursor (EC 3.4.23.34).
 CC GN CTSE.
 CC OS Cavia porcellus (Guinea pig).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
 CC OX NCBI_TaxID=10141;
 CC RN [1]
 CC RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 CC RX MEDLINE=92355614; PubMed=1644829;
 CC RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koiwai O.,
 CC RA Tanji M., Yakabe E., Athauda S.B., Takahashi K.;
 CC RT "Gastric procathepsin E and progastricsin from guinea pig.
 CC RT Purification, molecular cloning of cDNAs, and characterization of
 CC RT enzymatic properties, with special reference to procathepsin E.";
 CC RL J. Biol. Chem. 267:16450-16459(1992).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Gastric mucosa;
 CC RX MEDLINE=96073637; PubMed=8540321;
 CC RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,
 CC RA Tanji M., Athauda S.B., Takahashi K.;
 CC RT "Isolation, characterization, and structure of procathepsin E and
 CC RT cathepsin E from the gastric mucosa of guinea pig.";
 CC RL Adv. Exp. Med. Biol. 362:211-221(1995).
 CC CC -1- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
 CC CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
 CC CC -1- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
 CC CC specificity.
 CC CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 CC CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M88653; AAA37052.1; -

```

DR EMBL: S80547; AAB35844.1; -.
DR PIR: A43356; A43356.
DR HSP: P00794; 4CMS.
DR MEROPS: A01.010; -.
DR InterPro: IPR001461; Aspartaseal.
DR InterPro: IPR001969; Aspartaseal_site.
DR Pfam: PF00026; asp; 1.
DR PRINTS: PR00792; PEPIN.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 53
FT CHAIN 54 391
FT ACT_SITE 92 92
FT ACT_SITE 276 276
FT DISULFID 56 56
FT DISULFID 105 110
FT DISULFID 267 271
FT DISULFID 309 346
FT CARBOHYD 86 86
SQ SEQUENCE 391 AA; 42132 MW; 78D216BF8CFCDABD CRC64;

Query Match 1.4%; Score 9; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FLLLLLVLL 12
DB 4 FLLLLLVLL 12

RESULT 12
PPIE_HUMAN
ID PPIE_HUMAN STANDARD; PRT; 700 AA.
AC P23469;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-epsilon).
DE PTPRE.
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Placenta;
RX MEDLINE=91006018; PubMed=2170109;
RA Krueger N.X., Streuli M., Saito H.;
RT "Structural diversity and evolution of human receptor-like protein tyrosine phosphatases."
RT EMBO J. 9:3241-3252(1990).
RL -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O -> protein tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: X54134; CAA38069.1; -.
DR PIR: S12053; S12053.
DR HSP: P18052; 1YFO.
DR Genew: HGNC:9669; PTPRE.
DR MIM: 600926; -.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; Tyr_PP.

```

```

DR Pfam: PF00102; Y_phosphatase; 2.
DR PRINTS: PR00700; PTPPHPTASE.
DR SMART: SM00194; PTPC; 2.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 19
FT CHAIN 20 700
FT DOMAIN 20 46
FT TRANSMEM 47 69
FT DOMAIN 70 700
FT DOMAIN 154 393
FT DOMAIN 394 700
FT ACT_SITE 335 335
FT ACT_SITE 630 630
FT CARBOHYD 23 23
FT CARBOHYD 30 30
SQ SEQUENCE 700 AA; 80641 MW; D096BCADCEA65708 CRC64;

Query Match 1.4%; Score 9; DB 1; Length 700;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLVLL 13
DB 56 LLLLLLVLL 64

RESULT 13
CADL_CHICK
ID CADL_CHICK STANDARD; PRT; 732 AA.
AC P33145;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE B-cadherin precursor (K-CAM protein) (Fragment).
GN K-CAM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92107987; PubMed=1763068;
RA Sorkin B.C., Gallin W.J., Edelman G.M., Cunningham B.A.;
RT "Genes for two calcium-dependent cell adhesion molecules have similar structures and are arranged in tandem in the chicken genome."
RT Proc. Natl. Acad. Sci. U.S.A. 88:11545-11549(1991).
RN [2]
RP SEQUENCE OF 7-732 FROM N.A.
RC TISSUE=Embryonic brain;
RX MEDLINE=91325083; PubMed=2026653;
RA Napolitano E.W., Venstrom K., Wheeler E.F., Reichardt L.F.;
RT "Molecular cloning and characterization of B-cadherin, a novel chick cadherin."
RT J. Cell Biol. 113:893-905(1991).
CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS. THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE SORTING OF HETEROGENEOUS CELL TYPES. B-CADHERIN MAY HAVE IMPORTANT FUNCTIONS IN NEUROGENESIS, IN AT LEAST SOME EPITHELIA, AND IN EMBRYOGENESIS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES.
CC -1- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@lsb-sib.ch).
CC -----

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

CC EMBL; M81894; AAA48929.1; -;
CC EMBL; X58518; CAA41408.1; -;
CC PIR; A41634; IJCHCB.
CC HSP; P09803; 1SUH.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR000233; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 5.
CC Pfam; PF01049; Cadherin_C_term; 1.
CC SMART; SM00112; CA; 4.
CC PROSITE; PS00232; CADHERIN.1; 3.
CC PROSITE; PS00268; CADHERIN.2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT PROPEP 1
FT CHAIN <1 6 POTENTIAL.
FT DOMAIN 7 732 B-CADHERIN.
FT TRANSMEM 555 580 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 581 732 POTENTIAL.
FT DOMAIN 6 114 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 115 227 CADHERIN 1.
FT DOMAIN 228 339 CADHERIN 2.
FT DOMAIN 340 443 CADHERIN 3.
FT DOMAIN 444 554 CADHERIN 4.
FT DOMAIN 555 702 CADHERIN 5.
FT CARBOHYD 137 137 SER-RICH.
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 414 414 M -> V (IN REF. 2).
SQ SEQUENCE 732 AA; 80613 MW; 091D59A6A16CBD45 CRC64;

Query Match 1.4%; Score 9; DB 1; Length 732;

Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LLLVLLLL 15

Db 569 LLLVLLLL 577

RESULT 14

SZ05_HUMAN STANDARD; PRT; 114 AA.
AC P42830; Q960E1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Small inducible cytokine B5 precursor (CXCL5) (Epithelial-derived
DE neutrophil activating protein 78) (Neutrophil-activating peptide ENA-
DE 78).
DE SCYB5 OR ENA78.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95129887; PubMed=7828901;
RA Power C.A., Furness R.B., Brawand C., Wells T.N.C.;
RT "Cloning of a full-length cDNA encoding the neutrophil-activating
RT peptide ENA-78 from human platelets.";
RL Gene 151:333-334(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95014315; PubMed=7929219;
RA Chang M.S., McNinch J., Basu R., Simonet S.;
RT "Cloning and characterization of the human neutrophil-activating
RT peptide (ENA-78) gene.";
RL J. Biol. Chem. 269:25277-25282(1994).
RN [3]
RP SEQUENCE FROM N.A.

RX MEDLINE=95091791; PubMed=7999089;
RA Corbett M.S., Schmitt I., Riess O., Walz A.;
RT "Characterization of the gene for human neutrophil-activating peptide
RT 78 (ENA-78).";
RL Biochem. Biophys. Res. Commun. 205:612-617(1994).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=21361158; PubMed=11468158;
RA Zhang C., Thornton M.A., Kowalska M.A., Sachis B.S., Feldman M.,
RA Poncz M., McKenzie S.E., Reilly M.P.;
RT "Localization of distal regulatory domains in the
RT megakaryocyte-specific platelet basic protein/platelet factor 4 gene
RT locus.";
RL Blood 98:610-617(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RN Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 38-80 FROM N.A.
RA Anoli M.M., Thomson W., Hajeer A.H., Gonzalez-Gay M.A., Ollier W.E.R.;
RT "Novel polymorphism in ENA-78 gene";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 43-114 FROM N.A.
RC TISSUE=Epithelium;
RX MEDLINE=92078844; PubMed=1744577;
RA Walz A., Burgener R., Car B., Baggiolini M., Kunkel S.L.,
RA Strieter R.M.;
RT "Structure and neutrophil-activating properties of a novel
RT inflammatory peptide (ENA-78) with homology to interleukin 8.";
RL J. Exp. Med. 174:1355-1362(1991).
CC - FUNCTION: INVOLVED IN NEUTROPHIL ACTIVATION.
CC - SIMILARITY: BELONGS TO THE INTERCRINE ALPHA FAMILY (SMALL CYTOKINE
CC C-X-C) (CHEMOKINE CXCL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

CC EMBL; X78686; CAA53555.1; -;
CC EMBL; U12709; AAA62475.1; -;
CC EMBL; L37036; AAA86426.1; -;
CC EMBL; AJ315732; CAC42884.1; -;
CC EMBL; AF349466; AAK29641.1; -;
CC EMBL; BC008376; AAK08376.1; -;
CC HSP; P02775; 1NAP;
CC Genew; HGNC:10642; SCYB5.
CC MIM; 600324; -;
CC InterPro; IPR001089; CXCL_chmkine_smll.
CC InterPro; IPR001811; Chemokine_il8.
CC Pfam; PF00048; IL8; 1.
CC PRINTS; PR00437; SMALLCYTKCX.
CC SMART; SM00199; SCY; 1.
CC PROSITE; PS00471; SMALL_CYTOKINES_CXC; 1.
CC SIGNAL 1 36 POTENTIAL.
KW Cytokine; Signal.
FT CHAIN 37 114 SMALL INDUCIBLE CYTOKINE B5.
FT DISULFID 49 75 BY SIMILARITY.
FT DISULFID 51 91 BY SIMILARITY.
SQ SEQUENCE 114 AA; 11972 MW; 56B21EE86AE952D3 CRC64;

Query Match 1.2%; Score 8; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLLVLLLL 15
|||||||

Db 21 LLVLLLLL 28

RESULT 15

YPUD_BACSU STANDARD; PRT; 114 AA.

ID YPUD_BACSU STANDARD; PRT; 114 AA.

AC P17616;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein ypuD (ORF4).

GN ypuD.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / Marburg;

RA MEDLINE=95020538; PubMed=7934829;

RT "The organization of the Bacillus subtilis 168 chromosome region between the spoVA and serA genetic loci, based on sequence data.";

RL Mol. Microbiol. 10:385-395(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / SHGW;

RA Mironov V.N.;

RL Thesis (1989), USSR Academy of Sciences, Russia.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogawara N., Mosser I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita Y., Fujita Y., Funa S., Galizzi A., Galleron N., Chim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Guiseppi G., Guy B.J., Haga K., Haele J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogilwa A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";

RL Nature 390:249-256(1997).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; L09228; AAA67479.1; -.

DR EMBL; X51510; CAA35877.1; -.

DR EMBL; Z99116; CAB14262.1; -.

DR PIR; S45541; S45541.

DR Subtilist; BG10516; ypuD.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 114 AA; 12754 MW; 393036DFIA2DFC6D CRC64;

Query Match 1.2%; Score 8; DB 1; Length 114;

Best Local Similarity 100.0%; Pred. No. 3.6;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LLVLLLLL 15

DB 11 LLVLLLLL 18

Search completed: January 21, 2003, 09:57:48

Job time : 10.5326 secs

RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.,
RT "The DNA sequence of human chromosome 22.",
RL Nature 402:489-495(1999).
RN [5]
RP VARIANTS MLC L-93; R-118; R-212 AND L-280.
RX MEDLINE=21152271; PubMed=11254442;
RA Leegwater P.A.J., Yuan B.-Q., van der Steen J., Mulders J.,
RA Koenst A.A.M., Ilja Boor P.K., Mejaski-Bosnjak V.,
RA van der Maarel S.M., Frants R.R., Oudejans C.B.M., Schutgens R.B.H.,
RA Pronk J.C., van der Knapp M.S.,
RT "Mutations of MLC1 (KIAA0027), encoding a putative membrane protein,
RT cause megalencephalic leukoencephalopathy with subcortical cysts.",
RL Am. J. Hum. Genet. 68:831-838(2001).
CC -1- FUNCTION: May be a transporter. May act as a non-selective
CC neuronal cation channel.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: Expressed in the brain, with highest levels
CC found in the amygdala, nucleus caudatus, thalamus and hippocampus.
CC -1- DISEASE: Defects in MLC1 are a cause of megalencephalic
CC leukoencephalopathy with subcortical cysts (MLC). MLC is an
CC autosomal recessive disorder characterized by macrocephaly,
CC deterioration of motor functions with ataxia, and spasticity,
CC eventuating in mental decline. The brain appears swollen on
CC magnetic resonance imaging, with diffuse white-matter
CC abnormalities and the invariable presence of subcortical cysts.
CC -1- DISEASE: Defects in MLC1 are a cause of periodic catatonias (PC), a
CC familial subtype of catatonic schizophrenia which is a genetically
CC heterogeneous disorder characterized by psychosis and psychomotor
CC disturbances.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D25217; BAA04947.2; ALT_INIT.
DR EMBL; AF319633; AK660119.1;
DR EMBL; AL022327; CAB51559.1; ALT_INIT.
DR Genew; HGNC:17082; MLC1.
DR MIM; 605908;
DR MIM; 604004;
KW Transport; Transmembrane; Ionic channel; Disease mutation.
FT TRANSMEM 56 72 POTENTIAL.
FT TRANSMEM 113 128 POTENTIAL.
FT TRANSMEM 146 162 POTENTIAL.
FT TRANSMEM 200 216 POTENTIAL.
FT TRANSMEM 228 244 POTENTIAL.
FT TRANSMEM 302 318 POTENTIAL.
FT DOMAIN 262 280 PORE-REGION (POTENTIAL).
FT DOMAIN 147 156 POLY-LEU.
FT DOMAIN 173 176 POLY-LYS.
FT DOMAIN 305 315 POLY-LEU.
FT VARIANT 93 93
FT VARIANT 118 118
FT VARIANT 212 212
FT VARIANT 280 280
FT VARIANT 309 309
FT SEQUENCE 377 AA; 41141 MW; 9AF70B87D979F459 CRC64;
Query Match 1.2%; Score 9; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLLLLL 13
Db 306 LLLLLLLLLL 314
RESULT 6
CYB_EQUAS STANDARD; PRT; 379 AA.
AC P92487;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cytochrome B.
GN MTCYB OR COB OR CYTB.
OS Equus asinus (Donkey).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OC NCBI_TaxID=9793;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=97032591; PubMed=8875857;
RA Xu X., Gullberg A., Arnason U.,
RT "The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA
RT comparisons among four closely related mammalian species-pairs.",
RL J. Mol. Evol. 43:438-463(1996).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X97337; CAA66026.1;
DR HSP; P00157; IBE3.
DR InterPro; IPR000179; Cyt_b_b6.
DR Pfam; PF00032; cytochrome_b_c_1.
DR Pfam; PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Heme.
FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 182 182 IRON 2 (HEME B562 AXIAL LIGAND).
FT METAL 196 196 IRON 1 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 379 AA; 42791 MW; D4A6831CC486AA78 CRC64;
Query Match 1.2%; Score 9; DB 1; Length 379;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 LLLLLLLLLL 14
Db 232 LLLLLLLLLL 240
RESULT 7
CYB_ANTNA STANDARD; PRT; 381 AA.
ID CYB_ANTNA

Q33800;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytochrome B
 GN MTCYB OR COB OR CYTB
 OS Antechinus naso (Antechinus habbema).
 OG Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Antechinus.
 OX NCBI_TaxID=38774;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Armstrong L.A., Krajewski C., Westerman M.;
 RT "Phylogeny of the dasyurid marsupial genus Antechinus based on
 RL cytochrome b, 12S rRNA, and protamine P1 genes."; J. Mammal. 79:1379-1389(1998).
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS.
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN.
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U23461; AAB91425.1; -
 CC InterPro: IPR000179; Cyt_b_b6.
 CC Pfam: PF00032; cytochrome_b_c; 1.
 CC DR PROSITE; PS00193; CYTOCHROME_B_HEME; 1.
 CC KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
 CC Heme.
 CC METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
 CC FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
 CC FT METAL 182 182 IRON 2 (HEME B562 AXIAL LIGAND).
 CC FT METAL 196 196 IRON 1 (HEME B566 AXIAL LIGAND).
 CC SEQUENCE 381 AA; 42899 MW; 61C44C2BF7921EB0 CRC64;
 Query Match 1.2%; Score 9; DB 1; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 LLLLVLLLL 14
 Db 234 LLLLVLLLL 242
 RESULT 8
 ID CYB_DASHA STANDARD; PRT; 381 AA.
 AC Q34321;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Cytochrome B.
 GN MTCYB OR COB OR CYTB.
 OS Dasyurus hallucatus (Satanellus/northern quoll).
 OG Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Dasyurus.
 OX NCBI_TaxID=9280;
 RN [1]

SEQUENCE FROM N.A.
 RX MEDLINE=93096825; PubMed=1361058;
 RA Krajewski C., Driskell A.C., Baverstock P.R., Braun M.J.;
 RT "Phylogenetic relationships of the thylacine (Mammalia: Thylacinidae)
 RT among dasyurid marsupials: evidence from cytochrome b DNA
 RT sequences."; Proc. R. Soc. Lond., B, Biol. Sci. 250:19-27(1992).
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS.
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN.
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M99460; AAB40405.1; -
 CC InterPro: IPR000179; Cyt_b_b6.
 CC Pfam: PF00032; cytochrome_b_c; 1.
 CC DR PROSITE; PS00193; CYTOCHROME_B_HEME; 1.
 CC KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
 CC Heme.
 CC METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
 CC FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
 CC FT METAL 182 182 IRON 2 (HEME B562 AXIAL LIGAND).
 CC FT METAL 196 196 IRON 1 (HEME B566 AXIAL LIGAND).
 CC SEQUENCE 381 AA; 42914 MW; 49320F01FFABFB00 CRC64;
 Query Match 1.2%; Score 9; DB 1; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 LLLLVLLLL 14
 Db 234 LLLLVLLLL 242
 RESULT 9
 ID CYB_NEOLO STANDARD; PRT; 381 AA.
 AC Q35157;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Cytochrome B.
 GN MTCYB OR COB OR CYTB.
 OS Neophascogale lorentzii (Long-clawed marsupial mouse).
 OG Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Neophascogale.
 OX NCBI_TaxID=32551;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Krajewski C., Painter J., Buckley L., Westerman M.;
 RT "Phylogenetic structure of the marsupial family Dasyuridae based on
 RT cytochrome-b DNA sequences."; J. Mammal. Evol. 2:25-35(1994).
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 CC COUPLED TO ATP SYNTHESIS.
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY

CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U07585; AAB88761.1; -
CC InterPro: IPR001179; Cyt_b_b6.
CC Pfam: PF00032; cytochrome_b_c; 1.
CC PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
CC PROSITE: PS00193; CYTOCHROME_B_QO; 1.
CC Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
CC Heme.
CC METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
CC METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
CC METAL 182 182 IRON 1 (HEME B562 AXIAL LIGAND).
CC METAL 196 196 IRON 2 (HEME B566 AXIAL LIGAND).
CC SEQUENCE 381 AA; 42784 MW; 121C23A0B0B45A7E CRC64;
Query Match 1.2%; Score 9; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 LLLLVLLL 14
DB 234 LLLLVLLL 242
RESULT 10
CYB_PHATA STANDARD; PRT; 381 AA.
AC Q35673;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cytochrome B.
GN MTCYB OR COB OR CYTB.
OS Phasogale tapoatafa (Common wambenger).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Phascogale.
OX NCBI_TaxID=9293;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93096825; PubMed=1361058;
RA Krajewski C., Driskell A.C., Baverstock P.R., Braun M.J.;
RT "Phylogenetic relationships of the thylacine (Mammalia: Thylacinidae)
RT among dasyuroid marsupials: evidence from cytochrome b DNA
RT sequences";
RL Proc. R. Soc. Lond., B, Biol. Sci. 250:19-27(1992).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M99459; AAB40404.1; -
CC InterPro: IPR001179; Cyt_b_b6.
CC Pfam: PF00032; cytochrome_b_c; 1.
CC PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
CC PROSITE: PS00193; CYTOCHROME_B_QO; 1.
CC Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
CC Heme.
CC METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
CC METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
CC METAL 182 182 IRON 1 (HEME B562 AXIAL LIGAND).
CC METAL 196 196 IRON 2 (HEME B566 AXIAL LIGAND).
CC SEQUENCE 381 AA; 42755 MW; B8F15BEF4D5E41C7 CRC64;
Query Match 1.2%; Score 9; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 LLLLVLLL 14
DB 234 LLLLVLLL 242
RESULT 11
CATE_CAVPO STANDARD; PRT; 391 AA.
AC P25796;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cathepsin E precursor (EC 3.4.23.34).
GN CTSE.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=92355614; PubMed=1644829;
RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koizumi O.,
RA Tanji M., Yakabe E., Athauda S.B., Takahashi K.;
RT "Gastric procathepsin E and progastricsin from guinea pig.
RT Purification, molecular cloning of cDNAs, and characterization of
RT enzymatic properties, with special reference to procathepsin E";
RL J. Biol. Chem. 267:16450-16459(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Gastric mucosa;
RX MEDLINE=96073637; PubMed=8540321;
RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,
RA Tanji M., Athauda S.B., Takahashi K.;
RT "Isolation, characterization, and structure of procathepsin E and
RT cathepsin E from the gastric mucosa of guinea pig";
RL Adv. Exp. Med. Biol. 362:211-212(1995).
CC -1- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
CC -1- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
CC specificity.
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M88653; AAA37052.1; -
CC

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M99459; AAB40404.1; -
CC InterPro: IPR001179; Cyt_b_b6.
CC Pfam: PF00032; cytochrome_b_c; 1.
CC PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
CC PROSITE: PS00193; CYTOCHROME_B_QO; 1.
CC Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
CC Heme.
CC METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
CC METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
CC METAL 182 182 IRON 1 (HEME B562 AXIAL LIGAND).
CC METAL 196 196 IRON 2 (HEME B566 AXIAL LIGAND).
CC SEQUENCE 381 AA; 42755 MW; B8F15BEF4D5E41C7 CRC64;
Query Match 1.2%; Score 9; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 LLLLVLLL 14
DB 234 LLLLVLLL 242
RESULT 11
CATE_CAVPO STANDARD; PRT; 391 AA.
AC P25796;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cathepsin E precursor (EC 3.4.23.34).
GN CTSE.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=92355614; PubMed=1644829;
RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koizumi O.,
RA Tanji M., Yakabe E., Athauda S.B., Takahashi K.;
RT "Gastric procathepsin E and progastricsin from guinea pig.
RT Purification, molecular cloning of cDNAs, and characterization of
RT enzymatic properties, with special reference to procathepsin E";
RL J. Biol. Chem. 267:16450-16459(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Gastric mucosa;
RX MEDLINE=96073637; PubMed=8540321;
RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,
RA Tanji M., Athauda S.B., Takahashi K.;
RT "Isolation, characterization, and structure of procathepsin E and
RT cathepsin E from the gastric mucosa of guinea pig";
RL Adv. Exp. Med. Biol. 362:211-212(1995).
CC -1- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
CC -1- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
CC specificity.
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M88653; AAA37052.1; -
CC

DR EMBL; S80547; AAB35844.1; -
 DR PIR; A43356; A43356.
 DR HSSP; P00794; 4CMS.
 DR MEROPS; A01.010; -
 DR InterPro; IPR001461; AspartaseA1.
 DR InterPro; IPR001969; AspartaseA1.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PERSIN.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT PROPEP 19 53 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 54 391 CATHEPSIN E.
 FT ACT_SITE 92 92 BY SIMILARITY.
 FT ACT_SITE 276 276 BY SIMILARITY.
 FT DISULFID 56 56 INTERCHAIN (PROBABLE).
 FT DISULFID 105 110 BY SIMILARITY.
 FT DISULFID 267 271 BY SIMILARITY.
 FT DISULFID 309 346 BY SIMILARITY.
 FT CARBOHYD 86 86 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 391 AA; 42132 MW; 78D216BF8CFDABD CRC64;

Query Match 1.2%; Score 9; DB 1; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FLLLLLVLL 12
 |||||
 Db 4 FLLLLLVLL 12

RESULT 12
 ID PTPE_HUMAN STANDARD; PRT; 700 AA.
 AC P23469;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-epsilon).
 GN PTPRE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=91006018; PubMed=2170109;
 RA Krueger N.X., Streuli M., Saito H.;
 RT "Structural diversity and evolution of human receptor-like protein tyrosine phosphatases";
 RL EMBO J. 9:3241-3252(1990).
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O -> protein tyrosine + phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X54134; CAA38069.1; -
 DR PIR; S12053; S12053.
 DR HSSP; P18052; 1YFO.
 DR Genew; HGNC:9669; PTPRE.
 DR MIM; 600926; -
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; TYR_PP.

DR Pfam; PF00102; Y_phosphatase; 2.
 DR PRINTS; PR00700; PTPPHPTASE.
 DR SMART; SM00194; PTPC; 2.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
 DR PROSITE; PS00506; TYR_PHOSPHATASE_2; 2.
 DR PROSITE; PS00505; TYR_PHOSPHATASE_PTP; 2.
 KW Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat; Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 700 PROTEIN-TYROSINE PHOSPHATASE EPSILON.
 FT DOMAIN 20 46 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 47 69 POTENTIAL.
 FT DOMAIN 70 700 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 154 393 PROTEIN-TYROSINE PHOSPHATASE 1.
 FT DOMAIN 394 700 PROTEIN-TYROSINE PHOSPHATASE 2.
 FT ACT_SITE 335 335 BY SIMILARITY.
 FT ACT_SITE 630 630 BY SIMILARITY.
 FT CARBOHYD 23 23 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 30 30 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 700 AA; 80641 MW; D096BCADCEA65708 CRC64;

Query Match 1.2%; Score 9; DB 1; Length 700;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLVLL 13
 |||||
 Db 56 LLLLLLVLL 64

RESULT 13
 ID CADL_CHICK STANDARD; PRT; 732 AA.
 AC P31145;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE B-cadherin precursor (K-CAM protein) (fragment).
 GN K-CAM.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92107987; PubMed=1763068;
 RA Sorkin B.C., Gallin W.J., Edelman G.M., Cunningham B.A.;
 RT "Genes for two calcium-dependent cell adhesion molecules have similar structures and are arranged in tandem in the chicken genome.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:11545-11549(1991).
 RN [2]
 RP SEQUENCE OF 7-732 FROM N.A.
 RC TISSUE=Embryonic brain;
 RX MEDLINE=91225083; PubMed=2026653;
 RA Napolitano E.W., Venstrom K., Wheeler E.F., Reichardt L.F.;
 RT "Molecular cloning and characterization of B-cadherin, a novel chick cadherin";
 RL J. Cell Biol. 113:893-905(1991).
 CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS. THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE SORTING OF HETEROGENEOUS CELL TYPES. B-CADHERIN MAY HAVE IMPORTANT FUNCTIONS IN NEUROGENESIS, IN AT LEAST SOME EPITHELIA, AND IN EMBRYOGENESIS.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES.
 CC -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC MEDLINE-95091791; PubMed-7999089;
CC Corbett M.S., Schmitt I., Riess O., Walz A.;
CC "Characterization of the gene for human neutrophil-activating peptide
CC 78 (ENA-78).";
CC Biochem. Biophys. Res. Commun. 205:612-617(1994).
CC [4]
CC
CC SEQUENCE FROM N.A.
CC MEDLINE-21361158; PubMed-11468158;
CC Zhang C., Thornton M.A., Kowalska M.A., Sachis B.S., Feldman M.,
CC Poncz M., McKenzie S.E., Reilly M.P.;
CC "Localization of distal regulatory domains in the
CC megakaryocyte-specific platelet basic protein/platelet factor 4 gene
CC locus.";
CC Blood 98:610-617(2001).
CC [5]
CC SEQUENCE FROM N.A.
CC Strausberg R.;
CC TISSUE-Kidney;
CC Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC [6]
CC SEQUENCE OF 38-80 FROM N.A.
CC Anelli M.M., Thomson W., Hajeer A.H., Gonzalez-Gay M.A., Ollier W.E.R.;
CC "Novel polymorphism in ENA-78 gene.";
CC Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC [7]
CC SEQUENCE OF 43-114 FROM N.A.
CC TISSUE-Epithelium;
CC MEDLINE-92078844; PubMed-1744577;
CC Walz A., Burgener R., Car B., Baggolini M., Kunkel S.L.,
CC Strieter R.M.;
CC "Structure and neutrophil-activating properties of a novel
CC inflammatory peptide (ENA-78) with homology to interleukin 8.";
CC J. Exp. Med. 174:1355-1362(1991).
CC -1- FUNCTION: INVOLVED IN NEUTROPHIL ACTIVATION.
CC -1- SIMILARITY: BELONGS TO THE INTERCRINE ALPHA FAMILY (SMALL CYTOKINE
CC C-X-C) (CHEMOKINE CXCL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X78686; CAA55355.1; -
DR EMBL: U12709; AAA62475.1; -
DR EMBL: L37036; AAA86426.1; -
DR EMBL: AJ315732; CAC42884.1; -
DR EMBL: AF349466; AAK29641.1; -
DR EMBL: BC008376; AAH08376.1; -
DR HSSP: P02775; INAP. -
DR Genew; HGNC:10642; SCYB5.
DR MIM; 600324; -
DR InterPro; IPR001089; CXCL_chimkine_sm1.
DR InterPro; IPR001811; Chemokine_IL8.
DR Pfam; PF00048; IL8; 1.
DR PRINTS; PR00437; SMALLCYTCKXC.
DR SMART; SM00199; SCY; 1.
DR PROSITE; PS00471; SMALL_CYTOKINES_CXC; 1.
DR KW Cytokine; Signal.
DR FT SIGNAL 1 36 POTENTIAL
FT CHAIN 37 114 SMALL INDUCIBLE CYTOKINE B5.
FT DISULFID 49 75 BY SIMILARITY.
FT DISULFID 51 91 BY SIMILARITY.
SQ SEQUENCE 114 AA; 11972 MW; 5621EB86AE952D3 CRC64;

Query Match 1.2%; Score 9; DB 1; Length 732;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LLLVLLLLL 15
D5 569 LLLVLLLLL 577

RESULT 14
S205_HUMAN STANDARD; PRT; 114 AA.
ID P42830; Q960E1;
AC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Small inducible cytokine B5 precursor (CXCL5) (Epithelial-derived
DE neutrophil activating protein 78) (Neutrophil-activating peptide ENA-
DE 78).
GN SCYB5 OR ENA78.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
CC SEQUENCE FROM N.A.
CC MEDLINE-95129887; PubMed-7828901;
CC Power C.A., Furness R.B., Brawand C., Wells T.N.C.;
CC "Cloning of a full-length cDNA encoding the neutrophil-activating
CC peptide ENA-78 from human platelets.";
CC Gene 151:333-334(1994).
CC [2]
CC SEQUENCE FROM N.A.
CC MEDLINE-95014315; PubMed-7929219;
CC Chang M.S., McNinch J., Basu R., Simonet S.;
CC "Cloning and characterization of the human neutrophil-activating
CC peptide (ENA-78) gene.";
CC J. Biol. Chem. 269:25277-25282(1994).
CC [3]
CC SEQUENCE FROM N.A.

RX MEDLINE-95091791; PubMed-7999089;
RA Corbett M.S., Schmitt I., Riess O., Walz A.;
RT "Characterization of the gene for human neutrophil-activating peptide
RT 78 (ENA-78).";
RL Biochem. Biophys. Res. Commun. 205:612-617(1994).
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE-21361158; PubMed-11468158;
RA Zhang C., Thornton M.A., Kowalska M.A., Sachis B.S., Feldman M.,
RA Poncz M., McKenzie S.E., Reilly M.P.;
RT "Localization of distal regulatory domains in the
RT megakaryocyte-specific platelet basic protein/platelet factor 4 gene
RT locus.";
RL Blood 98:610-617(2001).
RN [5]
RN SEQUENCE FROM N.A.
RP Strausberg R.;
RC TISSUE-Kidney;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RN SEQUENCE OF 38-80 FROM N.A.
RA Anelli M.M., Thomson W., Hajeer A.H., Gonzalez-Gay M.A., Ollier W.E.R.;
RT "Novel polymorphism in ENA-78 gene.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RN SEQUENCE OF 43-114 FROM N.A.
RC TISSUE-Epithelium;
RX MEDLINE-92078844; PubMed-1744577;
RA Walz A., Burgener R., Car B., Baggolini M., Kunkel S.L.,
RA Strieter R.M.;
RT "Structure and neutrophil-activating properties of a novel
RT inflammatory peptide (ENA-78) with homology to interleukin 8.";
RL J. Exp. Med. 174:1355-1362(1991).
CC -1- FUNCTION: INVOLVED IN NEUTROPHIL ACTIVATION.
CC -1- SIMILARITY: BELONGS TO THE INTERCRINE ALPHA FAMILY (SMALL CYTOKINE
CC C-X-C) (CHEMOKINE CXCL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X78686; CAA55355.1; -
DR EMBL: U12709; AAA62475.1; -
DR EMBL: L37036; AAA86426.1; -
DR EMBL: AJ315732; CAC42884.1; -
DR EMBL: AF349466; AAK29641.1; -
DR EMBL: BC008376; AAH08376.1; -
DR HSSP: P02775; INAP. -
DR Genew; HGNC:10642; SCYB5.
DR MIM; 600324; -
DR InterPro; IPR001089; CXCL_chimkine_sm1.
DR InterPro; IPR001811; Chemokine_IL8.
DR Pfam; PF00048; IL8; 1.
DR PRINTS; PR00437; SMALLCYTCKXC.
DR SMART; SM00199; SCY; 1.
DR PROSITE; PS00471; SMALL_CYTOKINES_CXC; 1.
DR KW Cytokine; Signal.
DR FT SIGNAL 1 36 POTENTIAL
FT CHAIN 37 114 SMALL INDUCIBLE CYTOKINE B5.
FT DISULFID 49 75 BY SIMILARITY.
FT DISULFID 51 91 BY SIMILARITY.
SQ SEQUENCE 114 AA; 11972 MW; 5621EB86AE952D3 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLLVLLLLL 15
D5 114 LLLLLLL 15
D5 114 LLLLLLL 15

```

Db      21  LVLVLLLL 28
RESULT 15
YPUD_BACSU
ID      YPUD_BACSU      STANDARD;      PRT;      114 AA.
AC      P17616;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Hypothetical protein ypuD (ORFX4).
GN      YPU.D.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=1423;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168 / Marburg;
RX      MEDLINE=95020538; PubMed=7934829;
RA      Sorokin A.V., Zumstein E., Azevedo V., Ehrlich S.D., Serror P.;
RT      "The organization of the Bacillus subtilis 168 chromosome region
RL      between the spoVA and serA genetic loci, based on sequence data.";
RN      Mol. Microbiol. 10:385-395(1993).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168.
RA      Mironov V.N.;
RN      Thesis (1989), USSR Academy of Sciences, Russia.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168;
RX      MEDLINE=98044033; PubMed=9384377;
RA      Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA      Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA      Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA      Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA      Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA      Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA      Entian K.D., Errington J., Fabret C., Ferrali E., Foulger D.,
RA      Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA      Chim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA      Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA      Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA      Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA      Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA      Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA      Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA      Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA      Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA      Farro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA      Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA      Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA      Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA      Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
RA      Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA      Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA      Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA      Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA      Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA      Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT      "The complete genome sequence of the Gram-positive bacterium Bacillus
RT      subtilis.";
RL      Nature 390:249-256(1997).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; L09228; AAA67479.1; -.

```

```

DR      EMBL; X51510; CAA35877.1; -.
DR      EMBL; Z99116; CAB14262.1; -.
DR      PIR; S45541; S45541.
KW      Subtilisin; BG10516; ypuD.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 114 AA; 12754 MW; 393036DF1A2DFC6D CRC64;

Query Match      1.1%; Score 8; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8  LVLVLLLL 15
        |||||
DB      11  LVLVLLLL 18

Search completed: January 21, 2003, 09:57:46
Job time : 12.4674 secs

```

```

RESULT 13
O9XA18 PRELIMINARY; PRT; 362 AA.
ID Q9XA18;
AC Q9XA18;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative aspartate-semialdehyde dehydrogenase.
GN ASD OR SC03614 OR SC66T3.25C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL079348; CAB45481.1;
DR HSP; P00353; IPRM.
DR InterPro; IPR000534; Semialdh_dh.
DR Pfam; PF01118; Semialdehyde_dh; 1.
DR TIGRFAMs; TIGR01296; asd_B; 1.
SQ SEQUENCE 362 AA; 37988 MW; E186DE1BF7EA8D74 CRC64;

Alignment Scores:
Pred. No.: 19.4 Length: 362
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.89% Indels: 0
DB: 16 Gaps: 0

US-10-060-830-6 (1-275) x O9XA18 (1-362)

OY 207 GTATCAGCAGCAGGAGGATGGG 230
Db 172 ValSerGlyAlaGlyArgAspGly 179
|||||
RESULT 14
O28941 PRELIMINARY; PRT; 453 AA.
ID O28941
AC O28941;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Glycerol-3-phosphate dehydrogenase (GLPA).
GN AFI328.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16- / DSM 4304 / ATCC 49358;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Kyrpides J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides J.D.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE001012; AAB89918.1;
DR TIGR; AFI328;
DR InterPro; IPR000927; D_aa_oxidase.
DR Pfam; PF01266; DAO; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 453 AA; 49791 MW; F5A0869AC7253893 CRC64;

Alignment Scores:
Pred. No.: 18.8 Length: 453
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.89% Indels: 0
DB: 17 Gaps: 0

US-10-060-830-6 (1-275) x O28941 (1-453)

OY 88 TCTCCAGACTGACAGTGTCTCT 111
Db 331 SerProGlyLeuThrAlaAlaPro 338
|||||
RESULT 15
Q8XNC9 PRELIMINARY; PRT; 477 AA.
ID Q8XNC9
AC Q8XNC9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein CPE0409.
GN CPE0409.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003186; BAB80115.1;
DR InterPro; IPR002453; Beta_tubulin.
DR InterPro; IPR000927; D_aa_oxidase.

```

DR InterPro: IPR000205; NAD_binding.
DR Pfam: PF01266; DAO; 1.
DR PROSITE: PS00228; TUBULIN_B_AUTOREG; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 477 AA; 52163 MW; .3EA34FAE690872EC CRC64;

Alignment Scores:
Pred. No.: 18.7 Length: 477
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.89% Indels: 0
DB: 16 Gaps: 0

US-10-060-830-6 (1-275) x Q8XNC9 (1-477)

QY 88 TCTCCAGGACTGACAGCTGCTCCT 111
DB 335 SerProGlyLeuThrAlaAlaPro 342
|||||

Search completed: January 21, 2003, 14:52:24
Job time : 31.3661 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:46:52 ; Search time 9.96063 Seconds
(without alignments)
1624.657 Million cell updates/sec

Title: US-10-060-830-6
Perfect score: 90
Sequence: 1 caacttcagttggtcagccc.....tttgaagatgatgtgcttt 275

Scoring table:
OLIGO Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438908

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10060830/runat.16012003.092703.1523/app_query.fasta.1.1109
-DB-Issued_Patents_AA -QFWT=fastan -SUFFIX=olin2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830@cgn.1.1.8 -runat.16012003.092703.1523 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DRV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOB=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2.6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	8.9	574	2	US-08-756-317-4
2	8	8.9	600	2	US-08-756-317-3
3	8	8.9	638	2	US-08-756-317-2
4	7	7.9	37	4	US-09-152-060-55
5	7	7.9	188	5	PCT-US94-14277-5
6	7	7.9	267	4	US-08-871-572B-7
7	7	7.8	406	4	US-09-134-001C-3544
8	7	7.8	489	2	US-08-794-795-7
9	7	7.8	489	4	US-09-249-200-7
10	7	7.9	505	4	US-09-509-902A-11
11	7	7.8	518	1	US-08-392-367B-2
12	7	7.8	518	3	US-08-893-467A-2

13	7	7.8	619	2	US-08-373-652-5
14	7	7.8	619	5	PCT-US95-16311-5
15	7	7.8	1408	1	US-08-612-521-2
16	6	6.7	9	2	US-08-417-174-113
17	6	6.7	9	2	US-08-417-174-116
18	6	6.7	9	4	US-09-267-439-113
19	6	6.7	9	4	US-09-267-439-116
20	6	6.7	11	3	US-08-904-446A-9
21	6	6.7	14	1	US-07-923-724-53
22	6	6.7	14	1	US-08-609-426A-53
23	6	6.7	14	2	US-08-374-652C-45
24	6	6.7	15	2	US-07-728-215-42
25	6	6.7	15	4	US-08-938-085A-48
26	6	6.7	16	3	US-08-460-576-12
27	6	6.7	17	2	US-08-676-279-4
28	6	6.7	18	4	US-09-292-225-6
29	6	6.7	19	3	US-08-792-832A-36
30	6	6.7	19	4	US-08-652-877-18
31	6	6.7	19	4	US-08-476-515A-18
32	6	6.7	22	6	5281520-45
33	6	6.7	23	2	US-08-652-450A-15
34	6	6.7	23	2	US-08-374-652C-41
35	6	6.7	23	4	US-08-734-607B-6
36	6	6.7	23	6	5281520-6
37	6	6.7	25	4	US-09-512-342-4
38	6	6.7	26	2	US-08-703-660-33
39	6	6.7	26	3	US-08-989-045-33
40	6	6.7	28	4	US-09-312-183A-24
41	6	6.7	33	1	US-08-066-325-96
42	6	6.7	34	4	US-08-734-607B-5
43	6	6.7	34	4	US-08-734-607B-13
44	6	6.7	34	6	5281520-5
45	6	6.7	36	4	US-09-227-357-137

ALIGNMENTS

RESULT 1
US-08-756-317-4
; Sequence 4, Application US/08756317
; Patent No. 5849894
; GENERAL INFORMATION:
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
; TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthesis
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX USA
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,317
; FILING DATE: 25-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,693
; FILING DATE: 29-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT:008
; TELECOMMUNICATION INFORMATION:

```
; TELEPHONE: (713) 787-1400
; TELEFAX: (713) 787-1440
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-756-317-4
Alignment Scores:
Pred. No.: 6.21 Length: 574
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.89% Indels: 0
DB: 2 Gaps: 0
US-10-060-830-6 (1-275) x US-08-756-317-4 (1-574)
QY 24 ACATCCACTTTCAGGCTACGGG 47
Db 472 ThrSerThrPheLysAlaThrGly 479
RESULT 2
US-08-756-317-3
; Sequence 3, Application US/08756317
; Patent No. 5849894
; GENERAL INFORMATION:
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
; TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,317
; FILING DATE: 25-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,693
; FILING DATE: 29-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT:008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1400
; TELEFAX: (713) 787-1440
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 600 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-756-317-3
Alignment Scores:
Pred. No.: 6.16 Length: 600
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.89% Indels: 0
DB: 2 Gaps: 0
US-10-060-830-6 (1-275) x US-08-756-317-2 (1-638)
QY 24 ACATCCACTTTCAGGCTACGGG 47
Db 536 ThrSerThrPheLysAlaThrGly 543
RESULT 4
US-09-152-060-55
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.89% Indels: 0
DB: 2 Gaps: 0
US-10-060-830-6 (1-275) x US-08-756-317-3 (1-600)
QY 24 ACATCCACTTTCAGGCTACGGG 47
Db 498 ThrSerThrPheLysAlaThrGly 505
RESULT 3
US-08-756-317-2
; Sequence 2, Application US/08756317
; Patent No. 5849894
; GENERAL INFORMATION:
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
; TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,317
; FILING DATE: 25-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,693
; FILING DATE: 29-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT:008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1400
; TELEFAX: (713) 787-1440
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-756-317-2
Alignment Scores:
Pred. No.: 6.1 Length: 638
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.89% Indels: 0
DB: 2 Gaps: 0
US-10-060-830-6 (1-275) x US-08-756-317-2 (1-638)
QY 24 ACATCCACTTTCAGGCTACGGG 47
Db 536 ThrSerThrPheLysAlaThrGly 543
RESULT 4
US-09-152-060-55
```

Sequence 55, Application US/09152060
 Patent No. 6448230
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: 28 Human Secreted Proteins
 FILE REFERENCE: P2003P1.US
 CURRENT APPLICATION NUMBER: US/09/152,060
 CURRENT FILING DATE: 1998-09-11
 EARLIER APPLICATION NUMBER: PCT/US98/04858
 EARLIER FILING DATE: 1998-03-12
 EARLIER APPLICATION NUMBER: 60/040,762
 EARLIER FILING DATE: 1997-03-14
 EARLIER APPLICATION NUMBER: 60/040,710
 EARLIER FILING DATE: 1997-03-14
 EARLIER APPLICATION NUMBER: 60/050,934
 EARLIER FILING DATE: 1997-05-30
 EARLIER APPLICATION NUMBER: 60/048,100
 EARLIER FILING DATE: 1997-05-30
 EARLIER APPLICATION NUMBER: 60/048,357
 EARLIER FILING DATE: 1997-05-30
 EARLIER APPLICATION NUMBER: 60/048,189
 EARLIER FILING DATE: 1997-05-30
 EARLIER APPLICATION NUMBER: 60/057,765
 EARLIER FILING DATE: 1997-09-05
 EARLIER APPLICATION NUMBER: 60/048,970
 EARLIER FILING DATE: 1997-06-06
 EARLIER APPLICATION NUMBER: 60/068,368
 EARLIER FILING DATE: 1997-12-19
 NUMBER OF SEQ ID NOS: 118
 SOFTWARE: PatentIn ver. 2.0
 SEQ ID NO 55
 LENGTH: 37
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-152-060-55

Alignment Scores:
 Pred. No.: 88.2 Length: 37
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.87% Indels: 0
 DB: 4 Gaps: 0

US-10-060-830-6 (1-275) x US-09-152-060-55 (1-37)

QY 210 ATACTCTGTGCTGCTGTG 190
 DB 19 IleLeuLeuValCysSerVal 25
 |||||

RESULT 5
 PCT-US94-14277-5
 Sequence 5, Application PCT/US9414277
 GENERAL INFORMATION:
 APPLICANT: Aguet, Michel
 APPLICANT: Bohni, Ruth
 APPLICANT: Hemmi, Silvio
 TITLE OF INVENTION: Receptor Subunit Polypeptides
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/14277

FILING DATE: 07-DEC-1994
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/164596
 FILING DATE: 09-DEC-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Love, Richard B.
 REGISTRATION NUMBER: 34,659
 REFERENCE/DOCKET NUMBER: 866PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-5530
 TELEFAX: 415/952-9881
 TELEX: 910/371-7168
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 188 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 PCT-US94-14277-5
 Alignment Scores:
 Pred. No.: 67.4 Length: 188
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.87% Indels: 0
 DB: 5 Gaps: 0
 US-10-060-830-6 (1-275) x PCT-US94-14277-5 (1-188)
 QY 112 GAGGAGCAGCTGCTGCTGTG 92
 DB 81 GluGluGlnLeuSerValLeu 87
 |||||
 RESULT 6
 US-08-871-572B-7
 Sequence 7, Application US/08871572B
 Patent No. 6287853
 GENERAL INFORMATION:
 APPLICANT: Pestka, Sidney
 APPLICANT: Kotenko, Serguei
 APPLICANT: Soh, Jaemog
 APPLICANT: Donnelly, Robert
 APPLICANT: Mariano, Thomas
 APPLICANT: Cook, Jeffrey
 APPLICANT: Emmanuel, Stuart
 APPLICANT: Schwartz, Barbara
 TITLE OF INVENTION: Accessory Factor for Interferon Gamma
 NUMBER OF SEQUENCES: 17
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Richard R. Muccino
 STREET: 758 Springfield Avenue
 CITY: Summit
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07901
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/871,572B
 FILING DATE: 9-JUNE-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Muccino, Richard R.
 REGISTRATION NUMBER: 32,538
 REFERENCE/DOCKET NUMBER: UMD1-011
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 273-4988
 TELEFAX: (908) 273-4679

```

; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 267 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-871-572B-7

Alignment Scores:
Pred. No.: 63.6 Length: 267
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-6 (1-275) x US-08-871-572B-7 (1-267)
QY 112 GAGGAGCAGCTCAGTCTG 92
Db 145 GUGUGInLeuSerValleu 151

RESULT 7
US-09-134-001C-3544
; Sequence 3544, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3544
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3544

Alignment Scores:
Pred. No.: 59.3 Length: 406
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-6 (1-275) x US-09-134-001C-3544 (1-406)
QY 111 TCAGCCAGGCCAGCTATGAT 131
Db 348 SerAlaGlnAlaGlnTyrAsp 354

RESULT 8
US-08-794-795-7
; Sequence 7, Application US/08794795
; Patent No. 5916766
; GENERAL INFORMATION:
; APPLICANT: Elshourlagy, Nabil
; APPLICANT: Adamou, John
; APPLICANT: Gross, Mitchell
; APPLICANT: Lysko, Paul
; TITLE OF INVENTION: Human Macro Scavenger Rec
; TITLE OF INVENTION: eptor
; NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/794,795
FILING DATE: 04-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: ATG50009P
FILING DATE: 22-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-794-795-7

Alignment Scores:
Pred. No.: 57.5 Length: 489
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-6 (1-275) x US-08-794-795-7 (1-489)
QY 141 GCTGGAGCCAGGCTTACCT 161
Db 252 AlaglyLysProGlyLeuPro 258

RESULT 9
US-09-249-200-7
; Sequence 7, Application US/09249200
; Patent No. 6197931
; GENERAL INFORMATION:
; APPLICANT: ELSHOURBAGY, NABIL
; APPLICANT: ADAMOU, JOHN
; APPLICANT: GROSS, MITCHELL
; APPLICANT: LYSKO, PAUL
; TITLE OF INVENTION: HUMAN MARCO SCAVENGER RECEPTOR
; NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
```

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/249,200
;; FILING DATE: 12-FEB-1999
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/794,795
;; FILING DATE: 04-FEB-1997
;; APPLICATION NUMBER: 60/017,699
;; FILING DATE: 23-MAY-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Prestia, Paul F
;; REGISTRATION NUMBER: 23,031
;; REFERENCE/DOCKET NUMBER: ATG-50009-1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 610-407-0700
;; TELEFAX: 610-407-0700
;; TELEX: 846169
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 489 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-249-200-7

Alignment Scores:
Pred. No.: 57.5 Length: 489
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78% Indels: 0
Gaps: 0

US-10-060-830-6 (1-275) x US-09-249-200-7 (1-489)

QY 141 GCTGGAGCCAGGCTACCT 161
Db 252 AlAGLyLysProGlyLeuPro 258

RESULT 10
US-09-509-902A-11
;; Sequence 11, Application US/09509902A
;; Patent No. 6387676
;; GENERAL INFORMATION:
;; APPLICANT: Virca, Duke
;; APPLICANT: Bird, Timothy A.
;; APPLICANT: Anderson, Dirk M.
;; APPLICANT: Marken, John S.
;; TITLE OF INVENTION: Human cDNAs Encoding Polypeptides Having Kinase Functions
;; FILE REFERENCE: 2877-US
;; CURRENT APPLICATION NUMBER: US/09/509,902A
;; CURRENT FILING DATE: 1999-08-03
;; NUMBER OF SEQ ID NOS: 16
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 11
;; LENGTH: 505
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-509-902A-11

Alignment Scores:
Pred. No.: 57.2 Length: 505
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
Gaps: 0

US-10-060-830-6 (1-275) x US-09-509-902A-11 (1-505)

QY 59 GCGAGTGTGTTCCCGTAGC 39
|||||

Db 490 GlyArgLeuValProArgSer 496

RESULT 11
US-08-392-367B-2
;; Sequence 2, Application US/08392367B
;; Patent No. 5691197
;; GENERAL INFORMATION:
;; APPLICANT: Tryggvason, Karl
;; APPLICANT: Elomaa, Outi
;; APPLICANT: Kangas, Maarit
;; TITLE OF INVENTION: An Insolated DNA Sequence For a
;; Patent No. 5691197
;; TITLE OF INVENTION: No. 5691197el Macrophage Receptor with
;; TITLE OF INVENTION: a Collagenous Domain and the
;; TITLE OF INVENTION: Polypeptide Chain Encoded by
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
;; ADDRESSEE: Minnich & McKee
;; STREET: 1100 Superior Avenue
;; CITY: Suite 700
;; STATE: Cleveland
;; COUNTRY: Ohio
;; ZIP: 44114-2518
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch,
;; MEDIUM TYPE: 720 Kb storable
;; COMPUTER: IBM PS/2, Model 35 SX
;; OPERATING SYSTEM: DOS 5.0
;; SOFTWARE: Word Perfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/392,367B
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Minnich, Richard J.
;; REGISTRATION NUMBER: 24,175
;; REFERENCE/DOCKET NUMBER: TRV 2 009
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (216) 861-5582
;; TELEFAX: (216) 241-1666
;; TELEX: (216) 980162
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 518 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: Single
;; TOPOLOGY: Linear
US-08-392-367B-2

Alignment Scores:
Pred. No.: 57 Length: 518
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78% Indels: 0
Gaps: 0

US-10-060-830-6 (1-275) x US-08-392-367B-2 (1-518)

QY 141 GCTGGAGCCAGGCTACCT 161
Db 281 AlAGLyLysProGlyLeuPro 287

RESULT 12
US-08-893-467A-2
;; Sequence 2, Application US/08893467A
;; Patent No. 6063901
;; GENERAL INFORMATION:
;; APPLICANT: Tryggvason, Karl
;; APPLICANT: Elomaa, Outi

APPLICANT: Kangas, Maarit
TITLE OF INVENTION: An Insolated DNA Sequence For a
Patent No. 6063901
TITLE OF INVENTION: No. 6063901el Macrophage Receptor with
TITLE OF INVENTION: A Collagenous Domain and the
TITLE OF INVENTION: Polypeptide Chain Encoded by
TITLE OF INVENTION: such a Sequence
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fay, Sharpe, Beall, Fagan,
ADDRESSEE: Minnich & McKee
STREET: 1100 Superior Avenue
STREET: Suite 700
CITY: Cleveland
STATE: Ohio
COUNTRY: U.S.A.
ZIP: 44114-2518
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 Inch,
MEDIUM TYPE: 720 KB storable
COMPUTER: IBM PS/2, Model 35 SX
OPERATING SYSTEM: DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893.467A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Minnich, Richard J.
REGISTRATION NUMBER: 24,175
REFERENCE/DOCKET NUMBER: TRV 2 009
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 861-5582
TELEFAX: (216) 241-1666
TELEX: (216) 980162
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 518 amino acids
TYPE: amino acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-893-467A-2

Alignment Scores:
Pred. No.: 57 Length: 518
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78% Indels: 0
DB: 3 Gaps: 0

US-10-060-830-6 (1-275) x US-08-893-467A-2 (1-518)

QY 141 GCTGGGAGCCAGGCTACCT 161
Db 281 AlaGlyLysProGlyLeuPro 287

RESULT 13
US-08-372-652-5
Sequence 5, Application US/08372652
Patent No. 5932699
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Wongi
APPLICANT: Choi, Hueng-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA

COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/372.652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 619 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-372-652-5

Alignment Scores:
Pred. No.: 55.3 Length: 619
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-6 (1-275) x US-08-372-652-5 (1-619)

QY 15 CAGCCCTCCACATCCACTTTC 35
Db 255 GlnProSerThrSerThrPhe 261

RESULT 14
PCT-US95-16311-5
Sequence 5, Application PC/TUS9516311
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Wongi
APPLICANT: Choi, Hueng-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16311
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/372.652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001

TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 619 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US95-16311-5

Alignment Scores:
Pred. No.: 55.3 Length: 619
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78% Indels: 0
DB: 5 Gaps: 0

US-10-060-830-6 (1-275) x PCT-US95-16311-5 (1-619)

QY 15 CAGCGCTCCACATCCACTTC 35

Db 255 GinProSerThrSerThrPhe 261

RESULT 15

US-08-612-521-2
Sequence 2, Application US/08612521
Patent No. 5786483
GENERAL INFORMATION:
APPLICANT: Peery, Robert B
APPLICANT: Skatrud, Paul L
APPLICANT: Thornevell, Susan J
TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
TITLE OF INVENTION: CRYPTOCOCCUS NEOFORMANS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patent Division/AEH
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/612,521
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hamilton, Amy E
REGISTRATION NUMBER: 33,894
REFERENCE/DOCKET NUMBER: X-9693
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3169
TELEFAX: 317-276-1294
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1408 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-612-521-2

Alignment Scores:
Pred. No.: 48.3 Length: 1408
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x US-08-612-521-2 (1-1408)

QY 88 TCTCCAGGACTGACAGCTGCT 108

Db 4 SerProGlyLeuThrAlaAla 10

Search completed: January 21, 2003, 14:53:20
Job time : 10.9606 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:50:02 ; Search time 6.7126 seconds
(without alignments)
1628.756 Million cell updates/sec

Title: US-10-060-830-6

Perfect score: 90

Sequence: 1 caacttcagtgtgcagccc.....tttgaagatgatgtgcttt 275

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 120991 seqs, 19878514 residues

Word size: 1

Total number of hits satisfying chosen parameters: 218156

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10060830/runat_16012003_092704_1550/app_query.fasta_1.1109
-DB=PublishedApplications_AA -QFMT=fastan -SUFFIX=olin2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10060830 -CGN_1_1_4=runat_16012003_092704_1550
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

PublishedApplications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	9	10.1	505	10 US-09-910-689-307	Sequence 307, App
C 2	9	10.1	505	12 US-10-010-742-307	Sequence 307, App
C 3	8	9.0	685	10 US-09-801-574-6	Sequence 6, Appli
C 4	7	7.9	37	9 US-09-852-797-55	Sequence 55, Appl

c 5	7	7.9	37	10	US-09-853-161-55	Sequence 55, Appl
c 6	7	7.9	37	10	US-09-852-659A-55	Sequence 55, Appl
c 7	7	7.8	82	10	US-09-764-870-420	Sequence 420, App
c 8	7	7.9	88	9	US-09-796-692-1650	Sequence 1650, Ap
c 9	7	7.9	88	9	US-09-796-692-2501	Sequence 2501, Ap
c 10	7	7.8	213	9	US-09-738-826-4740	Sequence 4740, Ap
c 11	7	7.9	241	10	US-09-525-302-745	Sequence 745, App
c 12	7	7.8	302	10	US-09-764-853-590	Sequence 590, App
c 13	7	7.9	359	9	US-09-712-362-269	Sequence 269, App
c 14	7	7.9	396	9	US-10-047-542-799	Sequence 79, Appl
c 15	7	7.9	398	9	US-09-778-510-4	Sequence 4, Appli
c 16	7	7.9	495	10	US-09-815-242-5040	Sequence 5040, Ap
c 17	7	7.9	495	10	US-09-815-242-10315	Sequence 10315, A
c 18	7	7.9	495	10	US-09-815-242-11716	Sequence 11716, A
c 19	7	7.9	500	10	US-09-815-242-13764	Sequence 13764, A
c 20	7	7.9	505	9	US-09-931-457A-54	Sequence 54, Appl
c 21	7	7.8	509	9	US-10-028-072-108	Sequence 108, App
c 22	7	7.8	510	10	US-09-852-067-2	Sequence 2, Appli
c 23	7	7.9	724	9	US-09-965-529-32	Sequence 32, Appl
c 24	6	6.7	15	9	US-10-072-841-42	Sequence 42, Appl
c 25	6	6.7	15	10	US-09-947-770-22	Sequence 22, Appl
c 26	6	6.7	19	10	US-09-864-761-47914	Sequence 47914, A
c 27	6	6.7	20	10	US-09-864-761-43303	Sequence 43303, A
c 28	6	6.7	20	10	US-09-864-761-47432	Sequence 47432, A
c 29	6	6.7	38	10	US-09-864-761-46003	Sequence 46003, A
c 30	6	6.7	40	9	US-10-016-157A-210	Sequence 210, App
c 31	6	6.7	40	10	US-09-741-171-3	Sequence 3, Appli
c 32	6	6.7	43	10	US-09-864-761-35775	Sequence 35775, A
c 33	6	6.7	43	10	US-09-864-761-47198	Sequence 47198, A
c 34	6	6.7	44	10	US-09-864-761-41810	Sequence 41810, A
c 35	6	6.7	45	10	US-09-864-761-48841	Sequence 48841, A
c 36	6	6.7	46	9	US-10-059-749-13	Sequence 13, Appl
c 37	6	6.7	47	10	US-09-864-761-35186	Sequence 35186, A
c 38	6	6.7	47	10	US-09-864-761-33349	Sequence 33349, A
c 39	6	6.7	48	10	US-09-848-664-28	Sequence 28, Appl
c 40	6	6.7	50	9	US-09-796-692-1851	Sequence 1851, Ap
c 41	6	6.7	55	10	US-09-864-761-35633	Sequence 35633, A
c 42	6	6.7	55	10	US-09-864-761-42908	Sequence 42908, A
c 43	6	6.7	55	10	US-09-864-761-48670	Sequence 48670, A
c 44	6	6.7	55	10	US-09-864-761-47593	Sequence 47593, A
c 45	6	6.7	64	10	US-09-864-761-37333	Sequence 37333, A

ALIGNMENTS

RESULT 1
US-09-910-689-307
; Sequence 307, Application US/09910689
; Patent No. US20020081609A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Day, Craig H.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Wang, Tongtong
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.491C6
; CURRENT APPLICATION NUMBER: US/09/910,689
; CURRENT FILING DATE: 2001-07-20
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 307
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-910-689-307
Alignment Scores: 0.68
Pred. No.: 505

Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.11% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-6 (1-275) x US-09-910-689-307 (1-505)

QY 228 CATCCCTCTCTGCTCTGATCTTCT 202

Db 12 HisProPhLeuLeuLeuLeuLeu 20

RESULT 2

US-10-010-742-307

; Sequence 307, Application US/10010742

; Patent No. US20020146727A1

; GENERAL INFORMATION:

; APPLICANT: Dillon, Davin C.

; APPLICANT: Day, Craig H.

; APPLICANT: Jiang, Yuqiu

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Wang, Tongtong

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Bennington, Angela Ann

; APPLICANT: Zehentner, Barbara

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.491C7

; CURRENT APPLICATION NUMBER: US/10/010,742

; CURRENT FILING DATE: 2001-11-30

; NUMBER OF SEQ ID NOS: 307

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 307

; LENGTH: 505

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-010-742-307

Alignment Scores:

Pred. No.: 0.68

Score: 9.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 10.11%

DB: 12

Length: 505

Matches: 9

Conservatives: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-10-060-830-6 (1-275) x US-10-010-742-307 (1-505)

QY 228 CATCCCTCTCTGCTCTGATCTTCT 202

Db 12 HisProPhLeuLeuLeuLeuLeu 20

RESULT 3

US-09-801-574-6

; Sequence 6, Application US/09801574

; Patent No. US20020081592A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Peijiang Jeremy

; APPLICANT: Page, David C.

; TITLE OF INVENTION: Reproduction-Specific Genes

; FILE REFERENCE: 0399.2007-002

; CURRENT APPLICATION NUMBER: US/09/801,574

; CURRENT FILING DATE: 2001-03-07

; PRIOR APPLICATION NUMBER: 60/187,518

; PRIOR FILING DATE: 2000-03-07

; PRIOR APPLICATION NUMBER: 60/261,557

; PRIOR FILING DATE: 2001-01-12

; NUMBER OF SEQ ID NOS: 90

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6
; LENGTH: 685
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-801-574-6

Alignment Scores:

Pred. No.: 5.65

Score: 8.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 8.99%

DB: 10

Length: 685

Matches: 8

Conservatives: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-10-060-830-6 (1-275) x US-09-801-574-6 (1-685)

QY 226 TCCCTTCCTGCTCTGATCTTCT 203

Db 618 SerLeuProAlaProAspThrSer 625

RESULT 4

US-09-852-797-55

; Sequence 55, Application US/09852797

; Patent No. US20020172994A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; FILE REFERENCE: P2003P2

; TITLE OF INVENTION: 28 Human Secreted Proteins

; CURRENT APPLICATION NUMBER: US/09/852,797

; CURRENT FILING DATE: 2001-05-11

; PRIOR APPLICATION NUMBER: 60/265,583

; PRIOR FILING DATE: 2001-02-02

; PRIOR APPLICATION NUMBER: 09/152,060

; PRIOR FILING DATE: 1998-09-11

; PRIOR APPLICATION NUMBER: PCT/US98/04858

; PRIOR FILING DATE: 1998-03-12

; PRIOR APPLICATION NUMBER: 60/040,762

; PRIOR FILING DATE: 1997-03-14

; PRIOR APPLICATION NUMBER: 60/040,710

; PRIOR FILING DATE: 1997-03-14

; PRIOR APPLICATION NUMBER: 60/050,934

; PRIOR FILING DATE: 1997-05-30

; PRIOR APPLICATION NUMBER: 60/048,100

; PRIOR FILING DATE: 1997-05-30

; PRIOR APPLICATION NUMBER: 60/048,357

; PRIOR FILING DATE: 1997-05-30

; PRIOR APPLICATION NUMBER: 60/048,189

; PRIOR FILING DATE: 1997-05-30

; PRIOR APPLICATION NUMBER: 60/057,765

; PRIOR FILING DATE: 1997-09-05

; PRIOR APPLICATION NUMBER: 60/048,970

; PRIOR FILING DATE: 1997-06-06

; PRIOR APPLICATION NUMBER: 60/068,368

; PRIOR FILING DATE: 1997-12-19

; NUMBER OF SEQ ID NOS: 118

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 55

; LENGTH: 37

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-852-797-55

Alignment Scores:

Pred. No.: 73.4

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 7.87%

DB: 9

Length: 37

Matches: 7

Conservatives: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-10-060-830-6 (1-275) x US-09-852-797-55 (1-37)

QY 210 ATACTTCTGCTGCTCTGTG 190

```
Db 19 IleLeuLeuValCysSerVal 25
|||||
US-09-853-161-55
; Sequence 55, Application US/09853161
; Patent No. US20020076756A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PZ003P3
; CURRENT APPLICATION NUMBER: US/09/853,161
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 55
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-853-161-55

Alignment Scores:
Pred. No.: 73.4 Length: 37
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-6 (1-275) x US-09-853-161-55 (1-37)
QY 210 ATACTTCTGTGCTGTGTG 190
|||||
Db 19 IleLeuLeuValCysSerVal 25
|||||
RESULT 7
US-09-764-870-420
; Sequence 420, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ14
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 420
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-420

Alignment Scores:
Pred. No.: 65.8 Length: 82
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-6 (1-275) x US-09-764-870-420 (1-82)
QY 37 AGGCTACGGGAACCACTC 57
|||||
```

Db 58 ArgLeuArgGlyThrAsnLeu 64

RESULT 8

US-09-796-692-1650
 ; Sequence 1650, Application US/09796692
 ; Publication No. US20020198362A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gaiger, Alexander
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Mannion, Jane
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
 ; FILE REFERENCE: 2077.001200

US-10-060-830-6 (1-275) x US-09-796-692-1650 (1-88)

US-09-796-692-1650
 ; Sequence 1650, Application US/09796692
 ; Publication No. US20020198362A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gaiger, Alexander
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Mannion, Jane
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
 ; FILE REFERENCE: 2077.001200
 ; CURRENT APPLICATION NUMBER: US/09796.692
 ; CURRENT FILING DATE: 2001-03-01
 ; PRIOR APPLICATION NUMBER: 60/186,126
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: 60/190,479
 ; PRIOR FILING DATE: 2000-03-17
 ; PRIOR APPLICATION NUMBER: 60/200,545
 ; PRIOR FILING DATE: 2000-04-27
 ; PRIOR APPLICATION NUMBER: 60/200,303
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: 60/200,779
 ; PRIOR FILING DATE: 2000-05-04
 ; PRIOR APPLICATION NUMBER: 60/200,999
 ; PRIOR FILING DATE: 2000-05-22
 ; PRIOR APPLICATION NUMBER: 60/202,084
 ; PRIOR FILING DATE: 2000-07-14
 ; PRIOR APPLICATION NUMBER: 60/222,903
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: 60/223,416
 ; PRIOR FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: 60/223,378
 ; PRIOR FILING DATE: 2000-08-07
 ; NUMBER OF SEQ ID NOS: 9597
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1650
 ; LENGTH: 88
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

US-09-796-692-1650

US-09-796-692-1650

Alignment Scores:

Pred. No.: 65.1 Length: 88
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.87% Indels: 0
 DB: 9 Gaps: 0

US-10-060-830-6 (1-275) x US-09-796-692-1650 (1-88)

QY 169 TCTGGGCGAGTACCTGGC 149

Db 53 SerGlyAlaGlyArgProGly 59

RESULT 9

US-09-796-692-2501
 ; Sequence 2501, Application US/09796692
 ; Publication No. US20020198362A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gaiger, Alexander
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Mannion, Jane
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
 ; FILE REFERENCE: 2077.001200
 ; CURRENT APPLICATION NUMBER: US/09796.692

US-09-796-692-2501
 ; Sequence 2501, Application US/09796692
 ; Publication No. US20020198362A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gaiger, Alexander
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Mannion, Jane
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
 ; FILE REFERENCE: 2077.001200
 ; CURRENT APPLICATION NUMBER: US/09796.692
 ; CURRENT FILING DATE: 2001-03-01
 ; PRIOR APPLICATION NUMBER: 60/186,126
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: 60/190,479
 ; PRIOR FILING DATE: 2000-03-17
 ; PRIOR APPLICATION NUMBER: 60/200,545
 ; PRIOR FILING DATE: 2000-04-27
 ; PRIOR APPLICATION NUMBER: 60/200,303
 ; PRIOR FILING DATE: 2000-04-28
 ; PRIOR APPLICATION NUMBER: 60/200,779
 ; PRIOR FILING DATE: 2000-05-04
 ; PRIOR APPLICATION NUMBER: 60/200,999
 ; PRIOR FILING DATE: 2000-05-22
 ; PRIOR APPLICATION NUMBER: 60/202,084
 ; PRIOR FILING DATE: 2000-07-14
 ; PRIOR APPLICATION NUMBER: 60/222,903
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: 60/223,416
 ; PRIOR FILING DATE: 2000-08-04
 ; PRIOR APPLICATION NUMBER: 60/223,378
 ; PRIOR FILING DATE: 2000-08-07
 ; NUMBER OF SEQ ID NOS: 9597
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2501
 ; LENGTH: 88
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

US-09-796-692-2501

```
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4740
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4740

Alignment Scores:
Pred. No.: 57.6          Length: 213
Score: 7.00             Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78%      Indels: 0
DB: 9                   Gaps: 0

US-10-060-830-6 (1-275) x US-09-738-626-4740 (1-213)

Qy 153 GGTCTACTGCCCGACGAGAA 173
      |||||
Db 33 GlyLeuProAlaProAspGlu 39

RESULT 11
US-09-925-302-745
; Sequence 745, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 745
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (31)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (40)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-745

Alignment Scores:
Pred. No.: 56.6          Length: 241
Score: 7.00             Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87%      Indels: 0
DB: 10                  Gaps: 0

US-10-060-830-6 (1-275) x US-09-925-302-745 (1-241)

Qy 275 AAAGCAGCATCATCTCAAG 255
      |||||
Db 211 LysAlaAlaSerSerSerLys 217

RESULT 12
US-09-764-853-590
; Sequence 590, Application US/09764853
```

```
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P2206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 590
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (158)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (288)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-853-590

Alignment Scores:
Pred. No.: 54.9          Length: 302
Score: 7.00             Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.78%      Indels: 0
DB: 10                   Gaps: 0

US-10-060-830-6 (1-275) x US-09-764-853-590 (1-302)

Qy 81 ACACCTCTCTCCAGGACTGAC 101
      |||||
Db 168 ThrLeuLeuSerArgThrAsp 174

RESULT 13
US-09-712-363-269
; Sequence 269, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 269
; LENGTH: 359
; TYPE: PRT
```

; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-269

Alignment Scores: Length: 359
Pred. No.: 53.6
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 7.87%
DB: 9
Indels: 0
Gaps: 0

US-10-060-830-6 (1-275) x US-09-712-363-269 (1-359)

QY 110 GGAGCAGCTGTCAGTCTCGGA 90
 |||||
Db 258 GlyAlaLaValSerProGly 264

RESULT 14

US-10-047-542-79
; Sequence 79, Application US/10047542
; Patent No. US20020168367A1
; GENERAL INFORMATION:
; APPLICANT: WICOFF, JAMES W.
; TITLE OF INVENTION: NOVEL IMMUNOADHESINS FOR TREATING AND PREVENTING VIRAL
; FILE REFERENCE: 030905.0004.CIP1
; CURRENT APPLICATION NUMBER: US/10/047,542
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/US01/13932
; PRIOR FILING DATE: 2001-04-28
; PRIOR APPLICATION NUMBER: 60/200,298
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-047-542-79

Alignment Scores: Length: 396
Pred. No.: 52.8
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 7.87%
DB: 9
Indels: 0
Gaps: 0

US-10-060-830-6 (1-275) x US-10-047-542-79 (1-396)

QY 222 TTCCTGCTCCTGATCTCTT 202
 |||||
Db 340 PheLeuLeuLeuLeuLeu 346

RESULT 15

US-09-778-510-4
; Sequence 4, Application US/09778510
; Patent No. US20020164686A1
; GENERAL INFORMATION:
; APPLICANT: Baum, Peter
; TITLE OF INVENTION: Molecules Designated B7L1
; FILE REFERENCE: 2844-US
; CURRENT APPLICATION NUMBER: US/09/778,510
; CURRENT FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: PCT/US99/17906
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: 60/095,663
; PRIOR FILING DATE: 1998-08-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 398

; TYPE: PRT
; ORGANISM: Mus musculus
US-09-778-510-4
Alignment Scores: Length: 398
Pred. No.: 52.8
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 7.87%
DB: 9
Indels: 0
Gaps: 0

US-10-060-830-6 (1-275) x US-09-778-510-4 (1-398)

QY 222 TTCCTGCTCCTGATCTCTT 202
 |||||
Db 342 PheLeuLeuLeuLeuLeu 348

Search completed: January 21, 2003, 14:55:25
Job time : 6.7126 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:56:10 ; Search time 13.7026 seconds
(without alignments)
4581.302 Million cell updates/sec

Title: US-10-060-830-1114
Perfect score: 653
Sequence: 1 MPEFLLELVLLLELDAGA.....TQEVSGAGRGDCDFEKEIL 653

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	1.8	1807	2 JC6319	integrin beta-4 ch
2	10	1.5	546	2 B32688	beta-galactosidase
3	10	1.5	629	2 A46500	Ly-9.2 antigen - m
4	10	1.5	677	2 A32611	beta-galactosidase
5	9	1.4	276	2 G64584	conserved hypothet
6	9	1.4	379	2 T11375	ubiquinol-cytochro
7	9	1.4	391	2 A43356	cathepsin E (EC 3.
8	9	1.4	573	2 JC4335	anti-mullerian hor
9	9	1.4	700	1 S12053	protein-tyrosine-p
10	9	1.4	728	2 E83228	hypothetical prote
11	9	1.4	732	1 IJCHCB	B-cadherin precurs
12	9	1.4	1197	2 T30581	neural cell adhesi
13	8	1.2	95	2 C84649	hypothetical prote
14	8	1.2	95	2 S46537	pathogen-inducible
15	8	1.2	102	2 F87993	protein ZC334.3 [1
16	8	1.2	104	2 H90197	hypothetical prote
17	8	1.2	114	2 A55010	neutrophil-activat
18	8	1.2	114	2 S45541	hypothetical prote
19	8	1.2	125	2 T27519	hypothetical prote
20	8	1.2	161	2 T32037	hypothetical prote
21	8	1.2	201	2 H71059	hypothetical prote
22	8	1.2	222	2 AC2397	ATP-binding protei
23	8	1.2	236	2 T14169	hypothetical prote
24	8	1.2	258	2 C86541	Cri105 hypothetical
25	8	1.2	258	2 D72082	hypothetical prote
26	8	1.2	353	2 T07998	cytochrome c-type
27	8	1.2	353	2 T08001	cytochrome c-type
28	8	1.2	356	2 G01447	GP36b glycoprotein
29	8	1.2	358	2 JQ1278	histamine H2 recep

ALIGNMENTS

RESULT 1

JC6319
Integrin beta-4 chain precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 14-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jan-2000
C:Accession: JC6319
R:Feltri, M.L.; Arona, M.; Scherer, S.S.; Wraetz, L.
Gene 186, 299-304, 1997
A:Title: Cloning and sequence of the cDNA encoding the beta4 integrin subunit in rat
A:Reference number: JC6319; MUID:9728432; PMID:9074510
A:Accession: JC6319
A:Molecule type: mRNA
A:Residues: 1-1807 <PEL>
A:Cross-references: GB:U60096; NID:g1401302; PIDN:AAC53094.1; PID:g1401303
C:Experimental source: peripheral nerve
C:Comment: This protein belongs to a family of transmembrane receptors that mediate c
C:Superfamily: integrin beta-4 chain; EGF homology; fibronectin type III repeat homol
C:Keywords: glycoprotein; nerve; phosphoprotein; transmembrane protein
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-1807/Product: beta4 integrin #status predicted <MAT>
F:544-574/Domain: EGF homology <EGF>
F:713-735/Domain: transmembrane #status predicted <TM>
F:327,492,580,619,697/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1658/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 1.8%; Score 12; DB 2; Length 1807;
Best Local Similarity 100.0%; Pred. No. 0.0087;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 FLLLLLLLLLLL 15
| | | | | | | | | | | | | | | | |
Db 722 FLLLLLLLLLLL 733

RESULT 2

B32688
beta-galactosidase-related protein - human
C:Species: Homo sapiens (man)
C:Date: 29-Jun-1990 #sequence_revision 29-Jun-1990 #text_change 21-Jul-2000
C:Accession: B32688
R:Morreau, H.; Gajjar, N.J.; Gillemans, N.; Willemsen, R.; van der Horst, G.T.J.; d'
J. Biol. Chem. 264, 20655-20663, 1989
A:Title: Alternative splicing of beta-galactosidase mRNA generates the classic lysosc
A:Reference number: A32688; MUID:90062209; PMID:2511208
A:Accession: B32688
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-546 <MOR>
A:Cross-references: GB:M27508; NID:g179420; PIDN:AAA35599.1; PID:g179421
C:Keywords: alternative splicing

Query Match 1.5%; Score 10; DB 2; Length 546;

Best Local Similarity 100.0%; Pred. No. 0.29; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLLLL 14
 Db 9 LLLLLVLLLL 18

RESULT 3

A:Accession: A46500
 A:Molecule type: mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 18-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
 R:Sandrin, M.S.; Gumley, T.P.; Henning, M.M.; Vaughan, H.A.; Gones, L.J.; Trapani, J.A.; J. Immunol. 149, 1636-1641, 1992
 A:Title: Isolation and characterization of cDNA clones for mouse Ly-9.
 A:Reference number: A46500; MUID:92373005; PMID:1506686
 A:Accession: A46500
 A>Status: preliminary
 A:Molecule type: mRNA; protein
 A:Residues: 1-629 <SAS>
 A:Cross-references: GB:M84412; NID:gl198931; PIDN:AAA39468.1; PID:gl198932
 A:Experimental source: C57BL/6
 A:Note: sequence extracted from NCBI backbone (NCBIN:111651, NCBIP:111654)
 C:Keywords: transmembrane protein

Query Match 1.5%; Score 10; DB 2; Length 629;
 Best Local Similarity 100.0%; Pred. No. 0.33; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLLLL 14
 Db 435 LLLLLVLLLL 444

RESULT 4

A:Accession: A32611
 A:Molecule type: precursor - human
 N:Alternate names: lactase
 C:Species: Homo sapiens (man)
 C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 29-Sep-1999
 C:Accession: A32688; A32611; B37086; A31673
 R:Morreau, H.; Galjart, N.J.; Gillemans, N.; Willemsen, R.; van der Horst, G.T.J.; d'Azz
 J. Biol. Chem. 264, 20655-20663, 1989
 A:Title: Alternative splicing of beta-galactosidase mRNA generates the classic lysosomal
 A:Reference number: A32688; MUID:90062209; PMID:2511208

QY 5 LLLLLVLLLL 14
 Db 435 LLLLLVLLLL 444

A:Accession: A32688
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-677 <MOR>
 A:Cross-references: GB:M27507; NID:gl179400; PIDN:AAA51019.1; PID:gl179401
 R:Yamamoto, Y.; Hake, C.A.; Martin, B.M.; Kretz, K.A.; Ahern-Rindel, A.J.; Naylor, S.L.
 DNA Cell Biol. 9, 119-127, 1990
 A:Title: Isolation, characterization, and mapping of a human acid beta-galactosidase cDN
 A:Reference number: A32611; MUID:90262647; PMID:2111707
 A:Accession: A32611
 A:Molecule type: mRNA
 A:Residues: 1-677 <YAM>
 A:Cross-references: EMBL:M34423; NID:gl179422; PIDN:AAA51823.1; PID:gl179423
 A:Experimental source: testis
 R:Namba, E.; Suzuki, K.
 Biochem. Biophys. Res. Commun. 173, 141-148, 1990
 A:Title: Molecular cloning of mouse acid beta-galactosidase cDNA: sequence, expression c
 A:Reference number: A37086; MUID:91076843; PMID:2124109

A:Accession: B37086
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-677 <NAN>

R:Oshima, A.; Tsuji, A.; Nagao, Y.; Sakuraba, H.; Suzuki, Y.
 Biochem. Biophys. Res. Commun. 157, 238-244, 1988
 A:Title: Cloning, sequencing, and expression of cDNA for human beta-galactosidase.
 A:Reference number: A31673; MUID:89061717; PMID:3143362

A:Accession: A31673
 A:Molecule type: mRNA
 A:Residues: 1-9, P', 11-200, 'A', 202-677 <OSH>
 A:Cross-references: GB:M22590; NID:gl179418; PIDN:AAA51822.1; PID:gl179419
 A:Experimental source: placenta
 C:Comment: This enzyme is deficient in GM-1 gangliosidosis and Morquio B syndrome.
 C:Genetics:

A:Gene: GDB:GLE1
 A:Cross-references: GDB:119987; OMIM:230500
 A:Map position: 3p21.33-3p21.33
 C:Superfamily: beta-galactosidase bga
 C:Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-677/Product: beta-galactosidase #status predicted <MAT>
 F:26,247,464,496,542,545,555/Binding site: carbohydrate (Asn) (covalent) #status pred

Query Match 1.5%; Score 10; DB 2; Length 677;
 Best Local Similarity 100.0%; Pred. No. 0.35; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLLLL 14
 Db 9 LLLLLVLLLL 18

RESULT 5

G64584
 conserved hypothetical protein HP0519 - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C:Accession: G64584
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; Mcke
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
 Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467; PMID:9252185
 A:Accession: G64584
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-276 <TOM>
 A:Cross-references: GB:AE000566; GB:AE000511; NID:92313628; PIDN:AAD07587.1; PID:9231

Query Match 1.4%; Score 9; DB 2; Length 276;
 Best Local Similarity 100.0%; Pred. No. 1.6; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLLLL 13
 Db 16 LLLLLVLLLL 24

RESULT 6

T11375
 ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - donkey mitochondrion
 C:Species: mitochondrion Equus asinus (donkey)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
 C:Accession: T11375
 R:Xu, X.; Gullberg, A.; Arnason, U.
 J. Mol. Evol. 43, 438-463, 1996
 A:Title: The complete mitochondrial (mtDNA) of the donkey and mtDNA comparisons among
 A:Reference number: Z17265; MUID:97032591; PMID:8875857
 A:Accession: T11375
 A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-379 <XUX>
 A:Cross-references: EMBL:X97337; NID:gl1805746; PIDN:CAA66026.1; PID:gl1805759
 A:Experimental source: kidney
 C:Genetics:

A:Genome: mitochondrion
 C:Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastoqui
 C:Keywords: electron transfer; heme; iron; metalloprotein; mitochondrion; oxidative f

F;11-339/Domain: cytochrome b homology <CYB>
 F;11-209/Domain: cytochrome b6 homology <CB6>
 F;221-339/Domain: plastoquinol-plastocyanin reductase 17K protein homology <17K>
 F;83,182/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted
 F;97,196/Binding site: heme iron (His) (axial ligands) (high potential) #status predicted

Query Match 1.4%; Score 9; DB 2; Length 379;

Best Local Similarity 100.0%; Pred. No. 2.1;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LLLLLLLLLL 14

DB 232 LLLLLLLLLL 240

RESULT 7

A43356

cathepsin E (EC 3.4.23.34) precursor - guinea pig

N;Alternate names: erythrocyte membrane aspartic proteinase; slow-moving proteinase

C;Species: Cavia porcellus (guinea pig)

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999

C;Accession: A43356

R;Kageyama, T.; Ichinose, M.; Tsukada, S.; Miki, K.; Kurokawa, K.; Koiwai, O.; Tanji, M.

J. Biol. Chem. 267, 16450-16459, 1992

A;Title: Gastric procathepsin E and progastricsin from guinea pig. Purification, molecu

A;Reference number: A43356; MUID:92355614; PMID:1644829

A;Accession: A43356

A;Molecule type: mRNA

A;Residues: 1-391 <KAG>

A;CROSS-references: GB:M88653; NID:g191294; PIDN:AAA37052.1; PID:g191295

A;Note: sequence extracted from NCBI backbone (NCBIN:110763, NCBIPI:110769)

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; hydrolase; membrane protein

Query Match 1.4%; Score 9; DB 2; Length 391;

Best Local Similarity 100.0%; Pred. No. 2.1;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FLLLLLVLL 12

DB 4 FLLLLLVLL 12

RESULT 8

JC4335

anti-mullerian hormone type II receptor precursor - human

C;Species: Homo sapiens (man)

C;Date: 06-Dec-1995 #sequence_revision 08-Feb-1996 #text_change 04-Mar-2000

C;Accession: JC4335

R;Visser, J.A.; McLuskey, A.; van Beers, T.; Weghuis, D.O.; van Kessel, A.G.; Grootegeed

Biochem. Biophys. Res. Commun. 215, 1029-1036, 1995

A;Title: Structure and chromosomal localization of the human anti-muellerian hormone typ

A;Reference number: JC4335; MUID:96028015; PMID:7488027

A;Accession: JC4335

A;Molecule type: mRNA

A;Residues: 1-573 <VIS>

A;CROSS-references: GB:X91156; NID:g1107671; PIDN:CAA62593.1; PID:ei98046; PID:g1107672

C;Comment: This is a receptor for anti-mullerian hormone (see PIR:WFHUM). It plays a cri

C;Genetics:

A;Gene: GDB:AMHR2

A;CROSS-references: GDB:696210; OMIM:600956

A;Map position: 12q13-12q13

A;Introns: 16/3; 77/3; 141/3; 167/3; 207/2; 284/2; 322/3; 380/2; 429/3; 475/2

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C;Keywords: ATP; hormone receptor; transmembrane protein

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-573/Product: anti-muellerian hormone type II receptor #status predicted <MAT>

F;17-141/Domain: extracellular hormone binding #status predicted <ELB>

F;142-167/Domain: transmembrane #status predicted <TM>

F;201-512/Domain: protein kinase homology <KIN>

Query Match 1.4%; Score 9; DB 2; Length 573;

Best Local Similarity 100.0%; Pred. No. 3;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLLLLLLL 11

DB 154 LLLLLLLL 162

RESULT 9

SL2053

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type epsilon precursor - human

C;Species: Homo sapiens (man)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C;Accession: SL2053

R;Krueger, N.X.; Streuli, M.; Saito, H.

EMBO J. 9, 3241-3252, 1990

A;Title: Structural diversity and evolution of human receptor-like protein tyrosine p

A;Reference number: SL2049; MUID:9106018; PMID:2170109

A;Accession: SL2053

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-700 <KRU>

A;CROSS-references: GB:X54134; NID:g35791; PIDN:CAA38069.1; PID:g35792

C;Genetics:

A;Gene: GDB:PTPRE

A;CROSS-references: GDB:113185; OMIM:600926

A;Map position: 10q26-10q26

C;Function:

A;Description: catalyzes the hydrolysis of peptidyl-phosphotyrosine to release phosph

C;Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common an

C;Keywords: phosphoprotein; phosphoric monoester hydrolase; receptor; transmembrane p

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-700/Product: protein-tyrosine-phosphatase, receptor type epsilon #status predict

F;47-63/Domain: transmembrane #status predicted <TM>

F;78-698/Domain: leukocyte common antigen cytosolic domain homology <LAC>

F;159-383/Domain: protein-tyrosine-phosphatase homology <PTP>

F;335/Active site: Cys (phosphocysteine intermediate) #status predicted

F;341/Binding site: substrate phosphate (Arg) #status predicted

F;630/Active site: Cys (phosphocysteine intermediate) #status predicted

F;636/Binding site: substrate phosphate (Arg) #status predicted

Query Match 1.4%; Score 9; DB 1; Length 700;

Best Local Similarity 100.0%; Pred. No. 3.6;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLLL 13

DB 56 LLLLLLLL 64

RESULT 10

E83228

hypothetical protein PA3339 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C;Accession: E83228

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: E83228

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-728 <STO>

A;CROSS-references: GB:AE004756; GB:AE004091; NID:g9949466; PIDN:AA06727.1; GSPDB:GN

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA3339

Query Match 1.4%; Score 9; DB 2; Length 728;

Best Local Similarity 100.0%; Pred. No. 3.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology

Query Match 1.4% Score 9; DB 2; Length 1197;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LLLVLLLLL 14
| | | | | | | |
Db 1064 LLLVLLLLL 1072

RESULT 13
C84649
hypothetical protein At2g25510 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84649
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84649
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-95 <STO>
A:Cross-references: GB:AE002093; NID:g4432862; PIDN:AAD20710.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g25510
A:Map position: 2

Query Match 1.2% Score 8; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LFLVLLLV 10
| | | | | | | |
Db 7 LFLVLLLV 14

RESULT 14
S46537
pathogen-inducible protein CXC750 precursor - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jul-2000
C:Accession: S46537
R:Aufsatz, W.; Grimm, C.
Plant Mol. Biol. 25, 229-239, 1994
A:Title: A new, pathogen-inducible gene of Arabidopsis is expressed in an ecotype-spe
A:Reference number: S46537; MUID:94289647; PMID:8018872
A:Accession: S46537
A:Molecule type: DNA
A:Residues: 1-95 <AUF>
A:Cross-references: EMBL:X72022; NID:g457715; PIDN:CAA50905.1; PID:g457716
C:Genetics:
A:Gene: CXC750
A:Introns: 36/1
C:Function:
A:Description: probably a member of the plant defense response system
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-95/Product: pathogen-inducible protein CXC750 #status predicted <MAT>

Query Match 1.2% Score 8; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LFLVLLLV 10
| | | | | | | |
Db 11 LFLVLLLV 18

RESULT 15

QY 7 LLLVLLLLL 15
| | | | | | | |
Db 4 LLLVLLLLL 12

RESULT 11
IUCHCB
B-cadherin precursor - chicken (fragment)
N:Alternate names: K-CAM protein
C:Species: Gallus gallus (chicken)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: A1634; A38715; S16160
R:Sorkin, B.C.; Gallin, W.J.; Edelman, G.M.; Cunningham, B.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 11545-11549, 1991
A:Title: Genes for two calcium-dependent cell adhesion molecules have similar structures
A:Reference number: A41634; MUID:92107987; PMID:1763068
A:Accession: A41634
A:Molecule type: DNA
A:Residues: 1-732 <SOR>
A:Cross-references: GB:M81894; NID:g212226; PIDN:AAA48929.1; PID:g212227
R:Napolitano, E.W.; Venstrom, K.; Wheeler, E.F.; Reichardt, L.F.
J. Cell Biol. 113, 893-905, 1991
A:Title: Molecular cloning and characterization of B-cadherin, a novel chick cadherin.
A:Reference number: A38715; MUID:91225083; PMID:2026653
A:Accession: A38715
A:Molecule type: mRNA
A:Residues: 7-413, 'V', 415-732 <NAP>
A:Cross-references: GB:X58518; NID:g63113; PIDN:CAA41408.1; PID:g63114
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to
C:Genetics:

A:Gene: K-CAM
A:Introns: 29/3; 81/3; 130/1; 188/3; 231/3; 293/3; 375/2; 423/1; 498/1; 571/1; 614/3; 66
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pr
F:1-6/Domain: propeptide (fragment) #status predicted <PRO>
F:6-554/Domain: extracellular #status predicted <EXT>
F:7-732/Product: B-cadherin #status predicted <MAT>
F:9-114/Domain: cadherin repeat homology <CR1>
F:84-89/Region: cadherin binding #status predicted
F:117-227/Domain: cadherin repeat homology <CR2>
F:230-339/Domain: cadherin repeat homology <CR3>
F:340-447/Domain: cadherin repeat homology <CR4>
F:448-552/Domain: cadherin repeat homology <CR5>
F:555-580/Domain: transmembrane #status predicted <TM>
F:581-732/Domain: intracellular #status predicted <INT>
F:689-702/Region: serine-rich
F:137,410/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.4% Score 9; DB 1; Length 732;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LLLVLLLLL 15
| | | | | | | |
Db 569 LLLVLLLLL 577

RESULT 12
T30581
neural cell adhesion molecule L1.1 - zebra fish (fragment)
C:Species: Brachydanio rerio (zebra fish)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T30581
R:Tongliorgi, E.; Bernhardt, R.R.; Schachner, M.
J. Neurosci. Res. 42, 547-561, 1995
A:Title: Zebrafish neurons express two L1-related molecules during early axonogenesis.
A:Reference number: 220875; MUID:96155762; PMID:8568941
A:Accession: T30581
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1197 <TON>
A:Cross-references: EMBL:X89204; NID:g1065713; PID:g1065714; PIDN:CAA61490.1

F87993
 protein ZC334.3 [Imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C:Accession: F87993
 R:Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: F87993
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-102 <STO>
 A:Cross-references: GB:chr.I; PIDN:CAB04964.1; PID:g3881432; GSPDB:GN00019; CESP:ZC334.3
 C:Genetics:
 A:Note: predicted using GeneFinder
 A:Gene: ZC334.3
 A:Map position: 1

Query Match 1.2%; Score 8; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 6.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 LFLLLLLV 10
 |||||
 Db 28 LFLLLLLV 35

Search completed: January 21, 2003, 10:00:15
 Job time : 15.7026 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:54:45 ; Search time 8.03256 Seconds
(without alignments)
3371.781 Million cell updates/sec

Title: US-10-060-830-1114
Perfect score: 653
Sequence: 1 MFLFLLLVLLVLLLEDAGA.....TQEVSGAGRGDCVFKFEL 653

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	1.8	1807	1	Q64632 rattus norv
2	10	1.5	546	1	ITB4_RAT
3	10	1.5	654	1	Q01965 mus musculu
4	10	1.5	677	1	Q16278 homo sapien
5	9	1.4	377	1	Q15049 homo sapien
6	9	1.4	379	1	P92487 equus asinu
7	9	1.4	381	1	Q33800 antechinus
8	9	1.4	381	1	Q34321 dasyurus ha
9	9	1.4	381	1	Q35157 neophascoga
10	9	1.4	381	1	Q35673 phascogale
11	9	1.4	381	1	Q25796 cavia porce
12	9	1.4	700	1	P23469 homo sapien
13	9	1.4	732	1	P33145 gallus gall
14	8	1.2	114	1	P42830 homo sapien
15	8	1.2	114	1	P17616 bacillus su
16	8	1.2	242	1	Q8uk85 homo sapien
17	8	1.2	333	1	P48269 chlamydomon
18	8	1.2	356	1	P49256 canis fami
19	8	1.2	356	1	Q12907 homo sapien
20	8	1.2	358	1	P25102 rattus norv
21	8	1.2	381	1	Q34300 dasyuroides
22	8	1.2	381	1	Q34302 dasyuroides
23	8	1.2	396	1	P14091 homo sapien
24	8	1.2	397	1	P70269 mus musculu
25	8	1.2	425	1	P83088 drosophila
26	8	1.2	480	1	P10619 homo sapien
27	8	1.2	486	1	P39044 arabidopsis
28	8	1.2	491	1	P19535 bos taurus
29	8	1.2	576	1	P17644 drosophila
30	8	1.2	764	1	P47750 mus musculu
31	8	1.2	767	1	P36196 gallus gall
32	8	1.2	829	1	P22223 homo sapien
33	8	1.2	977	1	P09581 mus musculu

RESULT 1

ID	ITB4_RAT	STANDARD;	PRT;	1807 AA.
AC	Q64632;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Integrin beta-4 precursor (GP150) (CD104 antigen).			
GN	ITGB4.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=101116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-Sprague-Dawley; TISSUE-Sciatic nerve;			
RX	MEDLINE=9728432; Pubmed=9074510;			
RA	Reitri M.L., Arona M., Scherer S.S., Wrabetz L.;			
RT	"Cloning and sequence of the cDNA encoding the beta 4 Integrin subunit in rat peripheral nerve.";			
RL	Gene 186:299-304(1997).			
CC	-1- FUNCTION: INTEGRIN ALPHA-6/BETA-4 IS A RECEPTOR FOR LAMININ. IT PLAYS A CRITICAL STRUCTURAL ROLE IN THE HEMIDESMOSOME OF EPITHELIAL CELLS.			
CC	-1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-4 ASSOCIATES WITH ALPHA-6.			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- DOMAIN: THE FIBRONECTIN TYPE III-LIKE DOMAINS BIND BPAG1 AND PLECHIN AND PROBABLY ALSO RECRUIT BP230. CHAIN FAMILY.			
CC	-1- SIMILARITY: BELONGS TO THE INTEGRIN BETA TYPE III-LIKE DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U60096; AAC53094.1; .			
DR	HSP; P05106; LJV2.			
DR	InterPro; IPR003644; Calx_beta.			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR003961; FN_III.			
DR	InterPro; IPR003962; FNIII_repeat.			
DR	InterPro; IPR002369; Integrin_B.			
DR	InterPro; IPR001169; Integrin_beta_C.			
DR	InterPro; IPR003659; Plexin-like.			
DR	InterPro; IPR002035; VWF_A.			
DR	Pfam; PF00041; fn3_4.			
DR	Pfam; PF00362; integrin_B; 1.			
DR	Pfam; PF03160; Calx-beta; 1.			
DR	PRINTS; PR00014; FNTYPEIII.			
DR	PRODOM; PD001811; Integrin_B; 1.			

Q00495 rattus norv
Q9bz76 homo sapien
Q9c0a0 homo sapien
Q99p47 mus musculu
P01027 mus musculu
P01026 rattus norv
P19937 serratia ma
P69363 escherichia
P13345 escherichia
Q03709 escherichia
P10099 escherichia
P15176 escherichia

DR	SMART; SM00237; Calx_beta; 1.	GN	GLBI.
DR	SMART; SM00060; FN3; 4.	OS	Homo sapiens (Human).
DR	SMART; SM00187; INB; 1.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DR	SMART; SM00423; PSI; 1.	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
DR	SMART; SM00327; VWA; 1.	OX	NCBI_TaxID=9606;
DR	PROSITE; PS00243; INTEGRIN_BETA; 2.	RP	SEQUENCE FROM N.A., AND PLACENTAL, PARTIAL SEQUENCE.
DR	PROSITE; PS00022; EGF_1; UNKNOWN_2.	RC	TISSUE-Testis;
DR	PROSITE; PS01186; EGF_2; UNKNOWN_2.	RX	MEDLINE=90062209; PubMed=25111208;
DR	Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;	RA	Morreau H., Galjart N.J., Gillemaus N., Willemsen R.,
KW	Repeat; Signal.	RA	van der Horst G.T.J., D'Azzo A.;
FT	CHAIN 1 27	RA	"Alternative splicing of beta-galactosidase mRNA generates the
FT	CHAIN 28 1807	RT	classic lysosomal enzyme and a beta-galactosidase-related protein.";
FT	DOMAIN 28 712	RL	J. Biol. Chem. 264:20655-20663(1989).
FT	TRANSMEM 713 735	RN	[2]
FT	POTENTIAL 736 1807	RP	ELASTIN AND LAMININ BINDING DOMAIN.
FT	DOMAIN 736 1807	RX	PubMed=8383699;
FT	DOMAIN 129 366	RA	Hinek A., Rabinovitch M., Keeley F., Okamura-Oho Y., Callahan J.;
FT	DOMAIN 129 366	RT	"The 67-kD elastin/laminin-binding protein is related to an
FT	DOMAIN 457 503	RT	enzymatically inactive, alternatively spliced form of
FT	REPEAT 504 543	RT	beta-galactosidase.";
FT	REPEAT 544 582	RL	J. Clin. Invest. 91:1198-1205(1993).
FT	REPEAT 583 621	RN	[3]
FT	REPEAT 621 1128	RP	IDENTITY OF BETA-GALACTOSIDASE-RELATED PROTEIN WITH EBP.
FT	DOMAIN 1128 1216	RX	PubMed=9497360;
FT	DOMAIN 1221 1316	RA	Privitera S., Prody C.A., Callahan J.W., Hinek A.;
FT	DOMAIN 1458 1547	RT	"The 67-kDa enzymatically inactive alternative spliced variant of
FT	DOMAIN 1624 1716	RT	beta-galactosidase is identical to the elastin/laminin-binding
FT	DISULFID 30 456	RT	protein.";
FT	DISULFID 38 48	RL	J. Biol. Chem. 273:6319-6326(1998).
FT	DISULFID 41 72	RN	[4]
FT	DISULFID 51 61	RP	ELASTIC-FIBER ASSEMBLY STUDIES.
FT	DISULFID 245 288	RX	PubMed=10841810;
FT	DISULFID 424 673	RA	Hinek A., Zhang S., Smith A.C., Callahan J.W.;
FT	DISULFID 453 458	RT	"Impaired elastic-fiber assembly by fibroblasts from patients with
FT	DISULFID 469 480	RT	either Morquio B disease or infantile GM1-gangliosidosis is linked to
FT	DISULFID 477 513	RT	deficiency in the 67-kD spliced variant of beta-galactosidase.";
FT	DISULFID 482 491	RL	Am. J. Hum. Genet. 67:23-36(2000).
FT	DISULFID 493 504	RN	[5]
FT	DISULFID 519 524	RP	REVIEW.
FT	DISULFID 521 552	RX	PubMed=8922281;
FT	DISULFID 526 537	RA	Hinek A.;
FT	DISULFID 526 537	RT	"Biological roles of the non-integrin elastin/laminin receptor.";
FT	DISULFID 558 563	RL	Biol. Chem. 377:471-480(1996).
FT	DISULFID 565 574	RN	[6]
FT	DISULFID 576 583	RP	REVIEW.
FT	DISULFID 597 602	RX	PubMed=10571006;
FT	DISULFID 599 650	RA	Callahan J.W.;
FT	DISULFID 604 616	RT	"Molecular basis of GM1 gangliosidosis and Morquio disease, type B.
FT	DISULFID 628 637	RT	Structure-functional studies of lysosomal beta-galactosidase and the
FT	DISULFID 634 708	RT	non-lysosomal beta-galactosidase-like protein.";
FT	DISULFID 653 682	RT	Biochim. Biophys. Acta 1455:85-103(1999).
FT	CARBOHYD 327 327	CC	-1- FUNCTION: This protein has no beta-galactosidase catalytic
FT	CARBOHYD 492 492	CC	activity, but plays functional roles in the formation of
FT	CARBOHYD 580 580	CC	extracellular elastic fibers (elastogenesis) and in the
FT	CARBOHYD 619 619	CC	development of connective tissue. Seems to be identical to the
FT	CARBOHYD 697 697	CC	elastin-binding protein (EBP), a major component of the non-
SQ	SEQUENCE 1807 AA; 200588 MW; 2EF9BD4E345829A3 CRC64;	CC	integrin cell surface receptor expressed on fibroblasts, smooth
		CC	muscle cells, chondroblasts, leukocytes, and certain cancer cell
		CC	types. In elastin producing cells, associates with tropoelastin
		CC	intracellularly and functions as a recycling molecular chaperone
		CC	which facilitates the secretions of tropoelastin and its assembly
		CC	into elastic fibers.
		CC	-1- SUBCELLULAR LOCATION: Localized to the perinuclear area of the
		CC	cytoplasm but not to lysosomes.
		CC	-1- ALTERNATIVE PRODUCTS: 2 isoforms: 1/Beta-galactosidase (AC P162778)
		CC	and 2 (shown here); are produced by alternative splicing.
		CC	-1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
		CC	-----
		CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
		CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
		CC	the European Bioinformatics Institute. There are no restrictions on its
		CC	use by non-profit institutions as long as its content is in no way

Query Match 1.8%; Score 12; DB 1; Length 1807;
Best Local Similarity 100.08; Pred. No. 0.0049;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FULLLVLLLLL 15
| | | | | | | | | | | | | | | | | |
Db 722 FULLLVLLLLL 733

RESULT 2
ID BGAM_HUMAN STANDARD; PRT; 546 AA.
AC P16279;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Beta-galactosidase-related protein precursor (Beta-galactosidase-like
protein) (S-Gal) (Elastin-binding protein) (EBP).

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; M27508; AAA35599.1; -
 CC DR PUR; B32688; B32688.
 CC DR Genew; HGNC:4298; GLB1.
 CC DR MIM; 230500; -
 CC DR InterPro; IPR001944; GH_35.
 CC DR Pfam; PF01301; Glyco_hydro_35; 1.
 CC KW Signal; Alternative splicing; Glycoprotein.
 CC FT SIGNAL 1 23
 CC FT CHAIN 24 546 BETA-GALACTOSIDASE-RELATED PROTEIN.
 CC FT DOMAIN 90 103 ELASTIN/LAMININ BINDING..
 CC FT CARBOHYD 26 26 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SQ SEQUENCE 546 AA; 60551 MW; 2B1A73EDAF9E966C CRC64;

Query Match 1.5%; Score 10; DB 1; Length 546;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLLLLL 14
 DB 9 LLLLLLLLLL 18

RESULT 3

LY9_MOUSE STANDARD; PRT; 654 AA.
 ID LY9_MOUSE
 AC Q01965; O9ES35; O9ES36; O9ES29;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE T-lymphocyte surface antigen Ly-9 precursor (lymphocyte antigen 9)
 DE (Cell-surface molecule Ly-9).
 GN LY9 OR LY-9.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND POLYMORPHISM.
 RC STRAIN=129/SV, BALB/C, and C57BL/6; TISSUE=Spleen;
 RX MEDLINE=20424510; PubMed=10970093;
 RA Tovar V., de la Fuente M.A., Picueta P., Bosch J., Engel P.;
 RT "Gene structure of the mouse leukocyte cell surface molecule Ly9.";
 RL Immunogenetics 51:788-793(2000).
 RN [2]
 RP SEQUENCE OF 22-654 FROM N.A., AND SEQUENCE OF 48-59.
 RX MEDLINE=92373005; PubMed=1506686;
 RA Sandrin M.S., Gumley T.P., Henning M.M., Vaughan H.A., Gopez L.J.,
 RA Trapani J.A., McKenzie I.F.C.;
 RT "Isolation and characterization of cDNA clones for mouse Ly-9.";
 RL J. Immunol. 149:1636-1641(1992).
 CC -1- FUNCTION: May participate in adhesion reactions between T
 CC lymphocytes and accessory cells by homophilic interaction.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: LYMPHOCYTES.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; AF244131; AAG14997.1; -
 CC DR EMBL; AF244130; AAG14996.1; -
 CC DR EMBL; AF246701; AAG13268.2; -
 CC DR EMBL; AF245117; AAG13268.2; JOINED.
 CC DR EMBL; AF245506; AAG13268.2; JOINED.
 CC DR EMBL; AF245118; AAG13268.2; JOINED.
 CC DR EMBL; AF245507; AAG13268.2; JOINED.
 CC DR EMBL; AF245508; AAG13268.2; JOINED.
 CC DR EMBL; AF245509; AAG13268.2; JOINED.
 CC DR EMBL; AF245510; AAG13268.2; JOINED.
 CC DR EMBL; AF246699; AAG13268.2; JOINED.
 CC DR EMBL; AF246700; AAG13268.2; JOINED.
 CC DR EMBL; M84412; AAA39468.1; -
 CC DR PIR; A46500; A46500.
 CC DR HSP; P08921; LHNG.
 CC DR MGD; MGI:96885; Ly9.
 CC DR InterPro; IPR003599; Ig.
 CC DR InterPro; IPR003006; Ig_MHC.
 CC DR InterPro; IPR003600; Ig_like.
 CC DR Pfam; PF00047; Ig; 3.
 CC DR SMART; SM00409; Ig; 2.
 CC DR SMART; SM00410; Ig_like; 3.
 CC KW Cell adhesion; Antigen; Signal; Transmembrane; Glycoprotein;
 KW Immunoglobulin domain; Repeat; Polymorphism.
 CC FT SIGNAL 1 47
 CC FT CHAIN 48 654 T-LYMPHOCYTE SURFACE ANTIGEN LY-9.
 CC FT DOMAIN 48 453 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 454 474 POTENTIAL.
 CC FT DOMAIN 475 654 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 48 158 IG-LIKE V-TYPE DOMAIN 1.
 CC FT DOMAIN 163 249 IG-LIKE C2-TYPE DOMAIN 1.
 CC FT DOMAIN 250 362 IG-LIKE V-TYPE DOMAIN 2.
 CC FT DOMAIN 363 453 IG-LIKE C2-TYPE DOMAIN 2.
 CC FT DISULFID 172 242 POTENTIAL.
 CC FT DISULFID 178 222 POTENTIAL.
 CC FT DISULFID 376 445 POTENTIAL.
 CC FT DISULFID 382 426 POTENTIAL.
 CC FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 412 412 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT VARIANT 10 10 D -> G (IN LY9-1).
 CC FT VARIANT 14 14 G -> S (IN LY9-1).
 CC FT VARIANT 79 79 I -> T (IN LY9-1).
 CC FT VARIANT 91 91 F -> S (IN LY9-1).
 CC FT VARIANT 130 130 H -> Y (IN LY9-1).
 CC FT VARIANT 139 139 I -> T (IN LY9-1).
 CC FT VARIANT 362 362 P -> S
 CC FT VARIANT 366 366 K -> N (IN LY9-1).
 CC FT VARIANT 377 377 E -> K (IN LY9-1).
 CC FT VARIANT 550 550 M -> I (IN LY9-1).
 CC FT VARIANT 592 592 G -> E (IN LY9-1).
 CC FT CONFLICT 283 283 F -> L (IN REF. 2).
 CC FT CONFLICT 499 499 T -> P (IN REF. 2).
 CC FT CONFLICT 560 560 V -> L (IN REF. 2).
 CC FT CONFLICT 647 654 TPTYNET -> SPYL (IN REF. 2).
 CC SQ SEQUENCE 654 AA; 73142 MW; 1CBBE9708AE8E7 CRC64;

Query Match 1.5%; Score 10; DB 1; Length 654;
 Best Local Similarity 100.0%; Pred. No. 0.18;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLLLLL 14
 DB 456 LLLLLLLLLL 465

RT identification of four mutations in different clinical phenotypes
 RL among Japanese patients";
 Am. J. Hum. Genet. 49:566-574(1991).
 [9]
 RP VARIANTS GM1 THR-51; ARG-123; CYS-201; CYS-316 AND GLN-457.
 RX MEDLINE-91328151; PubMed-1907800;
 RA Yoshida K., Oshima A., Shimamoto M., Fukuhara Y., Sakuraba H.,
 RA Yanagisawa N., Suzuki Y.;
 RT "Human beta-galactosidase gene mutations in GM1-gangliosidosis: a
 RT common mutation among Japanese adult/chronic cases";
 Am. J. Hum. Genet. 49:435-442(1991).
 RL [10]
 RP VARIANT GM1 HIS-482.
 RX MEDLINE-93138608; PubMed-1487238;
 RA Mosna G., Fattore S., Tubiello G., Brocca S., Trubia M.,
 RA Gianazza E., Gatti R., Danesino C., Minelli A., Piantanida M.;
 RT "A homozygous missense arginine to histidine substitution at position
 RT 482 of the beta-galactosidase in an Italian infantile GM1-
 RT gangliosidosis patient";
 Am. J. Hum. Genet. 50:247-250(1992).
 RL [11]
 RP VARIANTS GM1 CYS-208; ARG-578; HIS-590 AND GLY-632.
 RX MEDLINE-94027054; PubMed-8213816;
 RA Boustany R.-M., Qian W.-H., Suzuki K.;
 RT "Mutations in acid beta-galactosidase cause GM1-gangliosidosis in
 RT American patients";
 Am. J. Hum. Genet. 53:881-888(1993).
 RL [12]
 RP VARIANT GM1 MET-82.
 RX MEDLINE-94256487; PubMed-8198123;
 RA Chakraborty S., Rafi M.A., Wenger D.A.;
 RT "Mutations in the lysosomal beta-galactosidase gene that cause the
 RT adult form of GM1 gangliosidosis";
 Am. J. Hum. Genet. 54:1004-1013(1994).
 RL [13]
 RP VARIANTS GM1 SER-148; TYR-214; ALA-216 AND GLY-532.
 RX Hilson W.L., Okamura-Oho Y., Zhang S., Clarke J.T.R., Mahuran D.,
 RA Callahan J.W.;
 RT "Novel missense mutations in beta-galactosidase that result in GM1-
 RT gangliosidosis";
 Am. J. Hum. Genet. 55:A223-A223(1994).
 RL [14]
 RP VARIANTS MORQUIO B HIS-83 AND CYS-482.
 RX MEDLINE-96049832; PubMed-7586649;
 RA Ishii N., Oohira T., Oshima A., Sakuraba H., Endo F., Matsuda I.,
 RA Sukeyama K., Orii T., Suzuki Y.;
 RT "Clinical and molecular analysis of a Japanese boy with Morquio B
 RT disease";
 Clin. Genet. 48:103-108(1995).
 RL [15]
 RP VARIANT GM1 SER-263.
 RX "Beta-galactosidase deficiency (beta-galactosidosis): GM1
 RT gangliosidosis and Morquio B disease";
 RL (in) Scriver C.R., Beaudet A.L., Sly W.S., Valle D. (eds.);
 RL The metabolic and molecular bases of inherited disease, pp.2787-2823,
 RL McGraw-Hill Publishing Co., New York (1995).
 RL [16]
 RP VARIANTS GM1 HIS-59; ASN-591 AND CYS-591.
 RX Morrone A., Bardelli T., Donati M.A., Giorgi M., Di Rocco R.,
 RA Gatti R., Taddeucci G., Ricci R., D'Azzo A., Zammarchi E.;
 RT "Identification of new mutations in six Italian patients affected by a
 RT variant form of infantile GM1-gangliosidosis with severe
 RT cardiomyopathy";
 Am. J. Hum. Genet. 61:A258-A258(1997).
 RL [17]
 RP VARIANTS GM1 HIS-201; SER-266 AND CYS-509.
 RX PubMed-9203065;
 RA Kaye E.M., Shalish C., Livermore J., Taylor H.A., Stevenson R.E.,
 RA Breakfield X.O.;
 RT "Beta-Galactosidase gene mutations in patients with slowly progressive
 RT GM1 gangliosidosis";
 J. Child Neurol. 12:242-247(1997).
 RL

RESULT 4
 BGAL_HUMAN STANDARD; PRT; 677 AA;
 AC P16278;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Beta-galactosidase precursor (EC 3.2.1.23) (Lactase) (Acid beta-
 DE galactosidase).
 DE GLB1.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-Testis;
 RX MEDLINE-90062209; PubMed-2511208;
 RA Morreau H., Galjart N.J., Gillemans N., Willemsen R.,
 RA van der Horst G.T.J., D'Azzo A.;
 RT "Alternative splicing of beta-galactosidase mRNA generates the
 RT classic lysosomal enzyme and a beta-galactosidase-related protein";
 RL J. Biol. Chem. 264:20655-20663(1989).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90262647; PubMed-2111707;
 RA Yamamoto Y., Hake C.A., Martin B.M., Kretz K.A., Ahern-Rindell A.J.,
 RA Naylor S.L., Mudd M., O'Brien J.S.;
 RT "Isolation, characterization, and mapping of a human acid beta-
 RT galactosidase cDNA";
 RL DNA Cell Biol. 9:119-127(1990).
 RL [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE-89061717; PubMed-3143362;
 RA Oshima A., Tsuji A., Nagao Y., Sakuraba H., Suzuki Y.;
 RT "Cloning, sequencing, and expression of cDNA for human beta-
 RT galactosidase";
 RL Biochem. Biophys. Res. Commun. 157:238-244(1988).
 RL [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Colon;
 RA Strausberg R.;
 RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RL [5]
 RP ELASTIC-FIBER ASSEMBLY STUDIES.
 RX PubMed-10841810;
 RA Hinek A., Zhang S., Smith A.C., Callahan J.W.;
 RT "Impaired elastic-fiber assembly by fibroblasts from patients with
 RT either Morquio B disease or infantile GM1-gangliosidosis is linked to
 RT deficiency in the 67-kD spliced variant of beta-galactosidase";
 RL Am. J. Hum. Genet. 67:23-36(2000).
 RL [6]
 RP REVIEW.
 RX PubMed-10571006;
 RA Callahan J.W.;
 RT "Molecular basis of GM1 gangliosidosis and Morquio disease, type B.
 RT Structure-function studies of lysosomal beta-galactosidase and the
 RT non-lysosomal beta-galactosidase-like protein";
 RL Biochim. Biophys. Acta 1455:85-103(1999).
 RL [7]
 RP VARIANTS MORQUIO B L-273; H-482 AND C-509, AND VARIANT GM1 C-494.
 RX MEDLINE-92026088; PubMed-1928092;
 RA Oshima A., Yoshida K., Shimamoto M., Fukuhara Y., Sakuraba H.,
 RA Suzuki Y.;
 RT "Human beta-galactosidase gene mutations in morquio B disease";
 RL Am. J. Hum. Genet. 49:1091-1093(1991).
 RL [8]
 RP VARIANTS GM1 CYS-49; THR-51 AND CYS-201.
 RX MEDLINE-91353572; PubMed-1909089;
 RA Nishimoto J., Nanba E., Inui K., Okada S., Suzuki K.;
 RT "GM1-gangliosidosis (genetic beta-galactosidase deficiency):

RN [18]
 RP VARIANTS MORQUITO B GLU-438; LYS-484 AND ALA-500.
 RA Skomrowski M.-A., Bigshaw R., Zhang S., Whelan D., Clarke J.T.R.,
 RA Callahan J.W.;
 RT "Novel mutations (Asn484Lys, Thr500Ala, Gly438Glu) in Morquio B
 RT disease";
 RL J. Inher. Metab. Dis. 22 Suppl. 1:131-131(1999).
 RN [19]
 RP VARIANTS GM1 H-59; S-121; C-208; M-240 AND N-491, AND VARIANTS P-10;
 RP C-521 AND G-532.
 RX MEDLINE-99268417; PubMed-10338095;
 RA Silva C.M.D., Severini M.H., Sopesa A., Coelho J.C., Zaha A.,
 RA d'Azzo A., Giugliani R.;
 RT "Six novel beta-galactosidase gene mutations in Brazilian patients
 RT with GM1-gangliosidosis";
 RL Hum. Mutat. 13:401-409(1999).
 RN [20]
 RP VARIANTS GM1 SER-148 AND ASN-332, AND VARIANT GLY-532.
 RP PubMed-10839995;
 RX Zhang S., Bagshaw R., Hilson W., Oho Y., Hinek A., Clarke J.T.R.,
 RA Hinek A., Callahan J.W.;
 RT "Characterization of beta-galactosidase mutations Asp332-->Asn and
 RT Arg148-->Ser, and a polymorphism, Ser532-->Gly, in a case of GM1
 RT gangliosidosis";
 RL Biochem. J. 348 3:621-632(2000).
 RN [21]
 RP VARIANTS GM1 HIS-59; HIS-201; HIS-482; ASP-579; ASN-591 AND CYS-591.
 RX PubMed-10737981;
 RA Morrone A., Bardelli T., Donati M.A., Giorgi M., Di Rocco M.,
 RA Gatti R., Parini R., Ricci R., Taddeucci G., D'Azzo A., Zammarchi E.;
 RT "Beta-galactosidase gene mutations affecting the lysosomal enzyme and
 RT the elastin-binding protein in GM1-gangliosidosis patients with
 RT cardiac involvement";
 RL Hum. Mutat. 15:354-366(2000).
 RN [22]
 RP VARIANT MORQUITO B L-273, AND VARIANTS GM1 M-82; H-201; D-270; Y-281;
 RP P-408 AND A-500.
 RX PubMed-11511921;
 RA Paschke E., Milos I., Kremler-Erlacher H., Hoefler G., Beck M.,
 RA Hoeltzenbein M., Kleijer W., Levade T., Michelakakis H., Radeva B.;
 RT "Mutation analyses in 17 patients with deficiency in acid
 RT beta-galactosidase: three novel point mutations and high correlation
 RT of mutation W273L with Morquio disease type B";
 RL Hum. Genet. 109:159-166(2001).
 CC -1- FUNCTION: Cleaves beta-linked terminal galactosyl residues from
 CC gangliosides, glycoproteins, and glycosaminoglycans.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
 CC galactose residues in beta-D-galactosides.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2/Beta-
 CC galactosidase-related protein (AC P16279); are produced by
 CC alternative splicing.
 CC -1- DISEASE: Defects in GLB1 are the cause of GM1-gangliosidosis, an
 CC autosomal recessive disorder with three major clinical phenotypes
 CC that are distinguished according to the age of onset and severity
 CC of symptoms: infantile (type I), juvenile (type II) and adult
 CC (type III). The infantile form is rapidly progressive and results
 CC in death within the first year of life.

Query Match 1.5%; Score 10; DB 1; Length 677;
 Best Local Similarity 100.0%; Pred. No. 0.19;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 LLLLVLLLLL 14
 Db 9 LLLLVLLLLL 18
 RESULT 5
 ID MLC1_HUMAN STANDARD; PRT; 377 AA.
 AC Q15049; Q5UGV8; Q96RP5;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Membrane protein MLC1
 GN MLC1 OR WCLI OR KIAA0027.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE-96051387; PubMed-7584026;
 RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
 RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
 RT "Prediction of the coding sequences of unidentified human genes. I.
 RT The coding sequences of 40 new genes (KIA0001-KIA0040) deduced by
 RT analysis of randomly sampled cDNA clones from human immature myeloid
 RT cell line KG-1";
 RL DNA Res. 1:27-35(1994).
 RN [2]
 RP REVISIONS.
 RA Ohara O., Nagase T., Kikuno R., Nomura N.;
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND VARIANT PC MET-309.
 RX TISSUE-Hippocampus;
 RC MEDLINE-21225990; PubMed-11326298;
 RA Meyer J., Huberth A., Ortega G., Syagailo Y.V., Jatzke S.,
 RA Moesner K., Strom T.M., Ulzheimer-Reuber I., Stoerber G., Schmitt A.,
 RA Lesch K.P.;
 RT "A missense mutation in a novel gene encoding a putative cation
 RT channel is associated with catatonic schizophrenia in a large
 RT pedigree";
 RL Mol. Psych. 6:302-306(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20057165; PubMed-10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Baggeley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn J.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
 RA McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odeil C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudo J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
 RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelman L.,

RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA Seroussi E., Fransson J., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkison P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.,
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
[5]
RN VARIANTS MLC L-93; R-118; R-212 AND L-280.
RP MEDLINE-21152271; PubMed-11254442.
RX Leeqwater P.A.J., Yuan B.Q., van der Steen J., Mulders J.,
RA Koest A.A.M., Ijla Boor P.K., Mejski-Bosnjak V., Schutgens R.B.H.,
RA van der Maarel S.M., Frants R.R., Oudejans C.B.M.,
RA Pronk J.C., van der Knapp M.S.;
RA "Mutations of MLC1 (KIA0027), encoding a putative membrane protein,
RT cause megalencephalic leukoencephalopathy with subcortical cysts.";
RL Am. J. Hum. Genet. 68:831-838(2001).
CC -1- FUNCTION: May be a transporter. May act as a non-selective
CC neuronal cation channel.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: Expressed in the brain, with highest levels
CC found in the amygdala, nucleus caudatus, thalamus and hippocampus.
CC -1- DISEASE: Defects in MLC1 are a cause of megalencephalic
CC leukoencephalopathy with subcortical cysts (MLC). MLC is an
CC autosomal recessive disorder characterized by macrocephaly,
CC deterioration of motor functions with ataxia, and spasticity,
CC eventuating in mental decline. The brain appears swollen on
CC magnetic resonance imaging, with diffuse white-matter
CC abnormalities and the invariable presence of subcortical cysts.
CC -1- DISEASE: Defects in MLC1 are a cause of periodic catatonias (PC), a
CC familial subtype of catatonic schizophrenia which is a genetically
CC heterogeneous disorder characterized by psychosis and psychomotor
CC disturbances.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D25217; BRA04947.2; ALT_INIT.
DR EMBL; AF319633; AAK60119.1; -.
DR EMBL; AL022327; CAB51559.1; ALT_INIT.
DR Genew; HGNC:17082; MLC1.
DR MIM; 605908; -.
DR MIM; 604004; -.
KW Transport; Transmembrane; Ionic channel; Disease mutation.
FT TRANSMEM 56 72
FT TRANSMEM 113 128
FT TRANSMEM 146 152
FT TRANSMEM 200 216
FT TRANSMEM 228 244
FT TRANSMEM 302 318
FT DOMAIN 262 280
FT DOMAIN 147 156
FT DOMAIN 173 176
FT DOMAIN 305 313
FT VARIANT 93 93
FT FTID-VAR_011699.
FT VARIAT 118 118
FT VARIAT 212 212
FT VARIAT 280 280
FT VARIAT 309 309
FT FTID-VAR_011700.
FT FTID-VAR_011701.
FT FTID-VAR_011702.
FT FTID-VAR_012731.
FT FTID-VAR_012731.
SQ SEQUENCE 377 AA; 41141 MW; 9AF70B87D979F459 CRC64;

Query Match 1.4%; Score 9; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 LLLLLLLLLL 13
DB 306 LLLLLLLLLL 314

RESULT 6
CYB_EQUAS STANDARD; PRT; 379 AA.
ID CYB_EQUAS
AC P92487;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cytochrome B,
GN MTCYB OR COB OR CYTB
OS Equus asinus (Donkey).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9793;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE-97032591; PubMed-8875857;
RA Xu X., Gullberg A., Arnason U.;
RT "The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA
RL J. Mol. Evol. 43:438-463(1996).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X97337; CRA66026.1; -.
DR HSSP; P00157; 1BE3.
DR InterPro; IPR000179; Cyt_b_b6.
DR Pfam; PF00032; cytochrome_b_c; 1.
DR Pfam; PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_QQ; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Heme.
FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 182 182 IRON 2 (HEME B562 AXIAL LIGAND).
FT METAL 196 196 IRON 1 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 379 AA; 42791 MW; D4A6831CC486AA78 CRC64;

Query Match 1.4%; Score 9; DB 1; Length 379;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
CYB_ANTNA STANDARD; PRT; 381 AA.
ID CYB_ANTNA

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 12:56:47 ; Search time 1381.93 Seconds
(without alignments)
3222.856 Million cell updates/sec

Title: US-10-060-830-6
Perfect score: 275
Sequence: 1 caacttcagtggtcagccc.....tttgaagatgatgctgcttt 275

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```
EST.*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estr:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_estfun:*
15: em_estom:*
16: gb_gss:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_othr:*
25: em_gss_pro:*
26: em_gss_rod:*
27: em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	275	100.0	647	9 AL705213	AL705213 DKFZp686A
2	224	81.5	680	14 BQ014739	BQ014739 UI-H-ED1-
3	190	69.1	1065	13 BM450083	BM450083 AGENCOURT
4	52	18.9	942	12 BF037277	BF037277 601460996
5	23	8.4	305	10 AW326693	AW326693 19588 MAR
6	23	8.4	345	12 BF543094	BF543094 UI-R-AFL-

7	23	8.4	447	9	AI596884	AI596884 vd98h08.y
8	23	8.4	459	9	AA467215	AA467215 vd98h08.r
c	9	8.4	479	10	BE098908	BE098908 UI-R-BJ1-
10	23	8.4	546	12	BG802724	BG802724 0184-46 M
11	23	8.4	1272	11	AK006805	AK006805 Mus muscu
12	21	7.6	298	9	AV204657	AV204657 AV204657
13	21	7.6	324	9	AV205839	AV205839 AV205839
14	21	7.6	326	9	AV047611	AV047611 AV047611
c	15	7.6	449	9	AI594861	AI594861 vd98h08.x
16	21	7.6	521	17	AQ930849	AQ930849 RPCI-23-2
c	17	7.6	680	14	BQ404474	BQ404474 GA_Ed007
18	21	7.6	690	10	BB327584	BB327584 BB327584
19	21	7.6	778	13	BG976310	BG976310 602846577
20	21	7.3	393	12	BF469974	BF469974 UI-M-BH3-
c	21	7.3	448	10	AM494758	AM494758 UI-M-BH3-
22	20	7.3	495	10	AV601798	AV601798 AV601798
c	23	7.3	513	12	BF369109	BF369109 RC3-GN007
c	24	7.3	542	10	AW491051	AW491051 UI-M-BH3-
25	20	7.3	912	12	BF167024	BF167024 601774157
c	26	6.9	224	14	Z36313	Z36313 MM208 MMV (
c	27	6.9	302	9	AI876295	AI876295 uk73e06.y
c	28	6.9	323	9	AI536909	AI536909 to14502.x
c	29	6.9	369	17	AZ070059	AZ070059 RPCI-23-4
c	30	6.9	448	13	BG923697	BG923697 602823205
c	31	6.9	489	9	AA120075	AA120075 mm23b12.1
c	32	6.9	483	12	BF746257	BF746257 RCL-BT025
c	33	6.9	483	12	BF746327	BF746327 RCL-BT025
c	34	6.9	493	14	W15477	W15477 zb05c05.r1
c	35	6.9	493	17	AZ069648	AZ069648 RPCI-23-4
c	36	6.9	494	12	BG632225	BG632225 GH05337.3
c	37	6.9	499	9	AU022203	AU022203 AU022203
c	38	6.9	499	10	BE108610	BE108610 UI-R-CAO-
c	39	6.9	526	13	BI320084	BI320084 lc45h07.y
c	40	6.9	527	17	AZ019281	AZ019281 RPCI-23-3
c	41	6.9	549	17	AQ993011	AQ993011 RPCI-23-3
c	42	6.9	566	10	BE571265	BE571265 601330702
c	43	6.9	586	13	BI792563	BI792563 lc38b01.y
c	44	6.9	590	17	AZ848249	AZ848249 2M0149H12
c	45	6.9	613	10	BE372299	BE372299 601223425

ALIGNMENTS

RESULT 1
AL705213
LOCUS
DEFINITION
DKFZp686A1235_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DKFZp686A1235 5', mRNA sequence.
AL705213
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 647)
Ottewaelder, B., Obermaier, B., Mewes, H.W., Weill, B. and
Wiemann, S.
EST (Ottewaelder, B., Obermaier, B., Mewes, H.W., Weill, B. and
Wiemann, S.)
TITLE
JOURNAL
COMMENT
Unpublished (2001)
Contact: Ottewaelder B
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by MedGenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. s1 sequence
also available.
This clone (DKFZp686A1235) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin- Charlottenburg, GERMANY; Email: clone@zpd.de.

FEATURES

source
1. .647
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686A1235"
/clone_lib="686 (synonym: hlcc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiI; cDNA-collection"

BASE COUNT 189 a 142 c 140 g 176 t

ORIGIN

Query Match 100.0%; Score 275; DB 9; Length 647;
Best Local Similarity 100.0%; Pred. No. 1.9e-135;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAACTTCAGTTGGTCAAGCCCTCCACATCCACTTCAAGGCTACGGGGAACCAACCTCCCC 60
|||||
Db 25 CAACTTCAGTTGGTCAAGCCCTCCACATCCACTTCAAGGCTACGGGGAACCAACCTCCCC 84
|||||
QY 61 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCTCAGCCAGG 120
|||||
Db 85 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCTCAGCCAGG 144
|||||
QY 121 CCAGTATGATACCCGGAAGCTGGGAAGCAGGTCTACTGCCCCAGACGAAATGGTGT 180
|||||
Db 145 CCAGTATGATACCCGGAAGCTGGGAAGCAGGTCTACTGCCCCAGACGAAATGGTGT 204
|||||
QY 181 ACCAGTGGCCAGAGACACACAGAAGTATCAGGACGAGGAGGATGGGAATGGATG 240
|||||
Db 205 ACCAGTGGCCAGAGACACACAGAAGTATCAGGACGAGGAGGATGGGAATGGATG 264
|||||
QY 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275
|||||
Db 265 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 299
|||||

RESULT 2
BQ014739/c 680 bp mRNA linear EST 26-MAR-2002
LOCUS
DEFINITION
UI-H-Ed1-act-h-12-0-UI-s1 NCI-CGAP_Ed1 Homo sapiens cDNA clone
IMAGE:5833547 3', mRNA sequence.

ACCESSION BQ014739
VERSION BQ014739.1 GI:19739640
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 680)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Seq primer: M13 FORWARD
POLVA=yes.

FEATURES

source
1. .680
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5833547"
/clone_lib="NCI-CGAP_Ed1"

/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Left Pubic Bone; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI-CGAP_Ed1 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma cell line C55. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996.
First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is GCTCAAGGCT.

TAG_LIB=UI-H-Ed1

TAG_TISSUE=chondrosarcoma

TAG_SEQ=CGTCAAGGCT

BASE COUNT 138 a 157 c 172 g 213 t
ORIGIN

Query Match 81.5%; Score 224; DB 14; Length 680;
Best Local Similarity 99.6%; Pred. No. 2.8e-108;
Matches 274; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAACTTCAGTTGGTCAAGCCCTCCACATCCACTTTCAGGCTAGCGGGAACCAACCTCCCC 60
|||||
Db 325 CAACTTCAGTTGGTCAAGCCCTCCACATCCACTTTCAGGCTAGCGGGAACCAACCTCCCC 266
|||||
QY 61 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGG 120
|||||
Db 265 CACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGG 206
|||||
QY 121 CCAGTATGATACCCGGAAGCTGGGAAGCAGGCTACCTGCCAGGAGGATGGGATGGTGT 180
|||||
Db 205 CCAGTATGATACCCGGAAGCTGGGAAGCAGGCTACCTGCCAGGAGGATGGGATGGTGT 146
|||||
QY 181 ACCAGTGGCCAGAGACACACAGAAGTATCAGGACGAGGAGGATGGGATGGATG 240
|||||
Db 145 ACCAGTGGCCAGAGACACACAGAAGTATCAGGACGAGGAGGATGGGATGGATG 86
|||||
QY 241 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 275
|||||
Db 85 TTTTAAAGAAATCCTTTGAAGATGATGCTGCTTT 51
|||||

RESULT 3
BQ450083 1065 bp mRNA linear EST 05-FEB-2002
LOCUS
DEFINITION
AGENCOURT_6393434 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528102
5', mRNA sequence.
ACCESSION BQ450083
VERSION BQ450083.1 GI:18499123
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1065)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; FILE REFERENCE: TUBERCULOSIS
 ; CURRENT APPLICATION NUMBER: US/09/103,840A
 ; CURRENT FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 4403765
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; FEATURE:
 ; OTHER INFORMATION: CDC 1551
 ; OTHER INFORMATION: "n" bases at various positions throughout the sequence
 ; OTHER INFORMATION: represent a, t, c or g
 ; US-09-103-840A-2

Query Match 6.2%; Score 17; DB 4; Length 4403765;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 TTCTCTCCAGGACTGAC 101
 |||||

Db 1712405 TTCTCTCCAGGACTGAC 1712389

RESULT 6
 US-09-363-939A-36
 ; Sequence 36, Application US/09363939A
 ; Patent No. 6346611
 ; GENERAL INFORMATION:
 ; APPLICANT: Pagratlis, Nikos
 ; APPLICANT: Lochrie, Michael
 ; APPLICANT: Gold, Larry
 ; TITLE OF INVENTION: High Affinity TGFbeta Nucleic Acid Ligands and
 ; TITLE OF INVENTION: Inhibitors
 ; FILE REFERENCE: NEX87
 ; CURRENT APPLICATION NUMBER: US/09/363,939A
 ; CURRENT FILING DATE: 1999-07-29
 ; PRIOR APPLICATION NUMBER: 09/046,247
 ; PRIOR FILING DATE: 1998-03-23
 ; PRIOR APPLICATION NUMBER: 08/458,424
 ; PRIOR FILING DATE: 1995-06-02
 ; PRIOR APPLICATION NUMBER: 07/714,131
 ; PRIOR FILING DATE: 1991-06-10
 ; PRIOR APPLICATION NUMBER: 07/931,473
 ; PRIOR FILING DATE: 1992-08-17
 ; PRIOR APPLICATION NUMBER: 07/964,624
 ; PRIOR FILING DATE: 1992-10-21
 ; PRIOR APPLICATION NUMBER: 08/117,991
 ; PRIOR FILING DATE: 1993-09-08
 ; PRIOR APPLICATION NUMBER: 07/536,428
 ; PRIOR FILING DATE: 1990-06-11
 ; NUMBER OF SEQ ID NOS: 216
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 36
 ; LENGTH: 71
 ; TYPE: RNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: Sequence
 ; NAME/KEY: modified_base
 ; LOCATION: (1)..(71)
 ; OTHER INFORMATION: All pyrimidines are 2'F.
 ; US-09-363-939A-36

Query Match 5.8%; Score 16; DB 4; Length 71;
 Best Local Similarity 87.5%; Pred. No. 29;
 Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 157 TACCTGCCCCGACGA 172
 Db 47 UACCGCCCCGACGA 62

RESULT 7

US-08-313-075A-41
 ; Sequence 41, Application US/08313075A
 ; Patent No. 5639870
 ; GENERAL INFORMATION:
 ; APPLICANT: Holton, Timothy A.
 ; APPLICANT: Cornish, Edwina C.
 ; APPLICANT: Tanaka, Yoshikazu
 ; TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOID
 ; TITLE OF INVENTION: PATHWAY ENZYMES AND USES THEREFOR
 ; NUMBER OF SEQUENCES: 58
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Scully, Scott, Murphy & Presser
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/313,075A
 ; FILING DATE: 30-NOV-1994
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: AU PL 1538/92
 ; FILING DATE: 27-MAR-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: AU PL 6598/93
 ; FILING DATE: 07-JAN-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: AU PCT/AU93/00127
 ; FILING DATE: 25-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Digiglio, Frank S.
 ; REGISTRATION NUMBER: 31,346
 ; REFERENCE/DOCKET NUMBER: 9433
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 742-4343
 ; TELEFAX: (516) 742-4366
 ; TELEX: 230 901 SANS UR
 ; INFORMATION FOR SEQ ID NO: 41:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 180 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA
 ; US-08-313-075A-41

Query Match 5.8%; Score 16; DB 1; Length 180;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 GATGTTTTTAAAGAAA 252
 Db 163 GATGTTTTTAAAGAAA 178

RESULT 8

US-08-589-939-8/c
 ; Sequence 8, Application US/08589939
 ; Patent No. 6015662
 ; GENERAL INFORMATION:
 ; APPLICANT: Hackett, Jr., John R.

APPLICANT: Hoff, Jane A.
APPLICANT: Ostrow, David H.
APPLICANT: Golden, Alan M.
TITLE OF INVENTION: REAGENTS FOR USE AS CALIBRATORS AND
TITLE OF INVENTION: CONTROLS
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/589,939
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 5865.US.01
TELEPHONE: 847-935-1729
TELEFAX: 847-938-2623
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 424 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-589-939-8

Query Match 5.8%; Score 16; DB 3; Length 424;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 TCTCCAGGACTGACAG 103
|||||
DB 119 TCTCCAGGACTGACAG 104

RESULT 9
US-08-647-368A-3/c
Sequence 3, Application US/08647368A
Patent No. 5928906
GENERAL INFORMATION:
APPLICANT: Koster, Hubert
APPLICANT: Van de Boom, Dirk
APPLICANT: Ruppert, Andreas
TITLE OF INVENTION: PROCESS FOR DIRECT SEQUENCING DURING
TITLE OF INVENTION: TEMPLATE AMPLIFICATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/647,368A
FILING DATE: 09-MAY-1996

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: SQA-020.01
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-647-368A-3

Query Match 5.8%; Score 16; DB 2; Length 558;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 AAAGAAATCCTTTGAA 261
|||||
DB 95 AAAGAAATCCTTTGAA 80

RESULT 10
US-09-328-111-656/c
Sequence 656, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Derti, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
TITLE OF INVENTION: PRODUCTS
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 656
LENGTH: 659
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(659)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-656

Query Match 5.8%; Score 16; DB 4; Length 659;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 TCAAGATGATGCTGCT 273
|||||
DB 445 TCAAGATGATGCTGCT 430

RESULT 11
US-07-791-931-1/c

```

; Sequence 1, Application US/07791931C
; Patent No. 6133507
; GENERAL INFORMATION:
; APPLICANT: Raikhel, Natasha V.
; TITLE OF INVENTION: Nettle Lectin cDNA
; FILE REFERENCE: MSU 4.1-114
; CURRENT APPLICATION NUMBER: US/07/791,931C
; CURRENT FILING DATE: 1991-11-12
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Urtica dioica
; US-07-791-931-1

Query Match      5.8%; Score 16; DB 3; Length 1317;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 134 CCCGAAAGCTGGGAAG 149
|||||
Db 534 CCCGAAAGCTGGGAAG 519

RESULT 12
US-09-443-184-44/c
; Sequence 44, Application US/09443184A
; Patent No. 6372431
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweiger, Gary
; APPLICANT: Kaser, Matthew R.
; APPLICANT: Panzer, Scott
; APPLICANT: Seilhammer, Jeffrey J.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah
; APPLICANT: Azimzai, Yalda
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS
; FILE REFERENCE: PC-0007 US
; CURRENT APPLICATION NUMBER: US/09/443,184A
; CURRENT FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 44
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6372431 1461451CB1
; US-09-443-184-44

Query Match      5.8%; Score 16; DB 4; Length 1376;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 CTGGGAAGCCAGGTCT 157
|||||
Db 825 CTGGGAAGCCAGGTCT 810

RESULT 13
US-09-227-357-18
; Sequence 18, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:49:12 ; Search time 14.7244 Seconds
(without alignments)
3590.903 Million cell updates/sec

Title: US-10-060-830-6
Perfect score: 90
Sequence: 1 caacttcagttggtcagccc.....tttgaagatgatgctgcttt 275

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565536

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US10080830/runat_16012003_092703_1513/app_query.fasta.1.1109
-DB=PIR_73 -QFMT=fastan -SUFFIX=oln2p.rpr -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10080830.ecgn.1.1.26.0runat_16012003_092703_1513 -NCPU=6 -ICPU=3
-NO_XLIFY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRAIDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	8.9	258	C86541	CT105 hypothetical
2	8	8.9	258	D72082	hypothetical prote
3	8	8.9	361	D72721	hypothetical prote
4	8	8.9	362	T35382	probable aspartate
5	8	9.0	366	S53073	hypothetical prote
6	8	8.9	384	S73429	glycerol-3-phosphat
7	8	8.9	453	G69415	conserved hypot
8	8	8.9	479	E72254	conserved hypot
9	8	9.0	512	B71474	probable hsp60 cha
10	8	8.9	826	AC0086	outer membrane ush
11	7	7.9	57	SHNZMV	small hydrophobic
12	7	7.9	57	SHNZME	small hydrophobic
13	7	7.9	57	SHNZMU	small hydrophobic
14	7	7.9	57	SHNZMS	small hydrophobic

c 15 7 7.9 57 1 SHNZMM
c 16 7 7.9 57 1 SHNZMT
c 17 7 7.9 57 1 SHNZBF
c 18 7 7.9 57 2 S24825
c 19 7 7.9 57 2 S24827
c 20 7 7.9 72 1 QQVZ9
c 21 7 7.9 102 2 S26409
c 22 7 7.9 105 2 E69058
c 23 7 7.9 113 2 S08455
c 24 7 7.9 119 2 S76274
c 25 7 7.9 146 2 S55861
c 26 7 7.9 158 2 T26692
c 27 7 7.9 172 2 F65038
c 28 7 7.9 172 2 B91062
c 29 7 7.9 172 2 F85906
c 30 7 7.9 197 2 A81113
c 31 7 7.8 200 2 T36305
c 32 7 7.8 206 2 H87328
c 33 7 7.9 222 2 A90099
c 34 7 7.9 222 2 B90128
c 35 7 7.8 232 2 E75547
c 36 7 7.8 241 2 T16849
c 37 7 7.8 250 2 T04968
c 38 7 7.9 251 2 G69076
c 39 7 7.8 255 2 S22825
c 40 7 7.8 255 2 S04105
c 41 7 7.8 255 2 JN0624
c 42 7 7.8 255 2 AC0023
c 43 7 7.9 266 2 E90354
c 44 7 7.9 270 2 E88646
c 45 7 7.9 278 2 S40929

ALIGNMENTS

RESULT 1

C86541

CT105 hypothetical protein [imported] - Chlamydothila pneumoniae (strain J138)
C:Species: Chlamydothila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: C86541
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: C86541
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <STO>
A:Cross-references: GB:BA000008; NID:g8978777; PIDN:BAA98613.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPj0405

Alignment Scores:
Pred. No.: 8.49 Length: 258
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.89% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-6 (1-275) x C86541 (1-258)

QY 6 TCATTTGGTCAGCCCTTCACATCC 29

|||||

Db 89 SerVaiglyGlnProSerThrSer 96

RESULT 2

D72082

hypothetical protein CP0350 [imported] - Chlamydothila pneumoniae (strains CWL029 an
N:Alternate names: ct105 hypothetical protein
C:Species: Chlamydothila pneumoniae, Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: D72082; C81587
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ollinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Accession: D72082

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-258 <ARN>

A:Cross-references: GB:AE001624; GB:AE001363; NID:g4376677; PIDN:AA18549.1; PID:g437668

A:Experimental source: strain CWL029

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,

C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: C81587

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-258 <REA>

A:Cross-references: GB:AE002197; GB:AE002161; NID:g7189269; PIDN:AAF38202.1; PID:g718927

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: CPn0405; CP0350

Alignment Scores:

Pred. No.:	8.49	Length:	258
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.89%	Indels:	0
DB:	2	Gaps:	0

US-10-060-830-6 (1-275) x D72082 (1-258)

QY 6 TCAGTTGGTCAGCGCTCCACATCC 29

|||||

Db 89 SerValGlyGlnProSerThrSer 96

RESULT 3

D72721

hypothetical protein APE0309 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999

C:Accession: D72721

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: D72721

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-361 <RAW>

A:Cross-references: DDBJ:AP000059; NID:g5103911; PIDN:BAA79264.1; PID:d1043050; PID:g510

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE0309

Alignment Scores:

Pred. No.:	8.23	Length:	361
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.89%	Indels:	0
DB:	2	Gaps:	0

US-10-060-830-6 (1-275) x D72721 (1-361)

QY 88 TCTCCAGGACTGACAGCTGCTCCT 111

|||||

Db 337 SerProGlyLeuThrAlaAlaPro 344

RESULT 4

T35382

probable aspartate-semialdehyde dehydrogenase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000

C:Accession: T35382

R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, June 1999

A:Reference number: Z21576

A:Accession: T35382

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-362 <MUR>

A:Cross-references: EMBL:AL079348; PIDN:CA845481.1; GSPDB:GN00070; SCOEDB:SC66T3.25c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: asd; SCOEDB:SC66T3.25c

C:Superfamily: aspartate-semialdehyde dehydrogenase

Alignment Scores:

Pred. No.:	8.23	Length:	362
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.89%	Indels:	0
DB:	2	Gaps:	0

US-10-060-830-6 (1-275) x T35382 (1-362)

QY 207 GTATCAGGACGAGGATGGG 230

|||||

Db 172 ValSerGlyAlaGlyArgAspGly 179

RESULT 5

S53073

hypothetical protein YMR251w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YMR251w

C:Species: Saccharomyces cerevisiae

C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C:Accession: S53073

R:Hunt, S.; Bowman, S.

submitted to the EMBL Data Library, March 1995

A:Reference number: S53069

A:Accession: S53073

A:Molecule type: DNA

A:Residues: 1-366 <HUN>

A:Cross-references: EMBL:Z48639; NID:g732924; PID:g732929; GSPDB:GN00013; MIPS:YMR25

C:Genetics:

A:Gene: MIPS:YMR251w

A:Cross-references: SGD:S0004863

A:Map position: 13R

Alignment Scores:

Pred. No.:	8.22	Length:	366
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.99%	Indels:	0
DB:	2	Gaps:	0

US-10-060-830-6 (1-275) x S53073 (1-366)

QY 56 AGTGTGTTCCCGTAGCCTTGAA 33

|||||

Db 194 ArgLeuValProArgSerLeuGlu 201

RESULT 6

S73429

glycerol-3-phosphate dehydrogenase glpD, aerob induced - Mycoplasma pneumoniae (strain

N:Alternate names: hypothetical protein D09_orf384

C:Species: Mycoplasma pneumoniae

A:Variety: ATCC 29342

C;Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 19-May-2000
C;Accession: S73429
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*
A;Reference number: S73327; MUID:97105885; PMID:8948633
A;Accession: S73429
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-384 <HIM>
A;Cross-references: EMBL:AE000012; GB:U00089; NID:g1673755; PIDN:AA895751.1; PID:g167375
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Genetics:
A;Gene: glpD
A;Genetic code: SGC3
C;Superfamily: *Escherichia coli* hypothetical protein ygaF

Alignment Scores:
Pred. No.: 8.18 Length: 384
Score: 8.00 Matches: 8
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 8.89%
Indels: 0
Gaps: 0
DB:

US-10-060-830-6 (1-275) x S73429 (1-384)
QY 88 TCTCCAGGACTGACAGCTGCTCCT 111
|||||
Db 348 SerProGlyLeuThrAlaAlaPro 355

RESULT 7
G69415
glycerol-3-phosphate dehydrogenase (glpA) homolog - *Archaeoglobus fulgidus*
C;Species: *Archaeoglobus fulgidus*
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: G69415
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: G69415
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-453 <KLE>
A;Cross-references: GB:AE001012; GB:AE000782; NID:g2689335; PIDN:AA889918.1; PID:g264924

Alignment Scores:
Pred. No.: 8.06 Length: 453
Score: 8.00 Matches: 8
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 8.89%
Indels: 0
Gaps: 0
DB:

US-10-060-830-6 (1-275) x G69415 (1-453)
QY 88 TCTCCAGGACTGACAGCTGCTCCT 111
|||||
Db 331 SerProGlyLeuThrAlaAlaPro 338

RESULT 8
E72254
conserved hypothetical protein - *Thermotoga maritima* (strain MSB8)
C;Species: *Thermotoga maritima*
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: E72254
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.

C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: E72254
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-479 <ARN>
A;Cross-references: GB:AE001795; GB:AE000512; NID:g4981992; PIDN:AA836502.1; PID:g49
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1432

Alignment Scores:
Pred. No.: 8.02 Length: 479
Score: 8.00 Matches: 8
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 8.89%
Indels: 0
Gaps: 0
DB:

US-10-060-830-6 (1-275) x E72254 (1-479)
QY 88 TCTCCAGGACTGACAGCTGCTCCT 111
|||||
Db 339 SerProGlyLeuThrAlaAlaPro 346

RESULT 9
B71474
probable hsp60 chaperonin, mitochondrial - *Chlamydia trachomatis* (serotype D, strain
C;Species: *Chlamydia trachomatis*
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 17-Mar-2000
C;Accession: B71474
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitch
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia*
A;Reference number: A71570; MUID:99000809; PMID:9784136
A;Accession: B71474
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <ARN>
A;Cross-references: GB:AE001348; GB:AE001273; NID:g3329216; PIDN:AA868350.1; PID:g33
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: C7755
C;Superfamily: chaperonin groEL
C;Keywords: mitochondrion

Alignment Scores:
Pred. No.: 7.97 Length: 512
Score: 8.00 Matches: 8
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 8.99%
Indels: 0
Gaps: 0
DB:

US-10-060-830-6 (1-275) x B71474 (1-512)
QY 266 TCATCTTCAAAGGATTTCTTTAAA 243
|||||
Db 37 SerSerSerLysAspPhePheLys 44

RESULT 10
AC0086
outer membrane usher protein YPO0698 [imported] - *Yersinia pestis* (strain CO92)
C;Species: *Yersinia pestis*
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AC0086
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice,
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan,
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barre
Nature 413, 523-527, 2001
A;Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0086
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-826 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89550.1; PID:g15978784; GSPDB:GN00175
C;Genetics:
A;Gene: YPO0698

Alignment Scores:
Pred. No.: 7.63 Length: 826
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.89% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-6 (1-275) x AC0086 (1-826)

QY 3 ACTTCAGTTGGTCAGCCTCCACA 26
|||||
Db 356 ThrSerValGlyGlnProSerThr 363

RESULT 11

SHNZMV
small hydrophobic protein - mumps virus (strains SBL-1 and SBL)

C;Species: mumps virus
C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jun-2000
C;Accession: A31884; S24823; C60004
R;Elango, N.; Koevamees, J.; Varsanyi, T.M.; Norrby, E.
J. Virol. 63, 1413-1415, 1989

A;Title: mRNA sequence and deduced amino acid sequence of the mumps virus small hydrophobic protein
A;Reference number: A31884; MUID:89125739; PMID:2915385
A;Accession: A31884
A;Molecule type: mRNA
A;Residues: 1-57 <ELA>
A;Cross-references: GB:M25421; NID:g556278; PIDN:AAA50290.1; PID:g556279
A;Experimental source: strain SBL-1
R;Yeo, R.P.; Afzal, M.A.; Forsey, T.; Rima, B.K.
submitted to the EMBL Data Library, January 1992
A;Description: Nucleotide sequence analysis of the SH gene of mumps virus reveals several
A;Reference number: S19866

Alignment Scores:
Pred. No.: 110 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x SHNZMV (1-57)
QY 222 TTCCTGCTCCTGATCTCTT 202
|||||
Db 12 PheLeuLeuLeuLeuLeu 18

RESULT 13
SHNZMU
small hydrophobic protein - mumps virus (strain Urabe Am9)
C;Species: mumps virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 07-May-1999
C;Accession: A38479; JU0308; JQ2366; S11637
R;Turner, P.C.; Forsey, T.; Minor, P.D.
J. Gen. Virol. 72, 435-437, 1991
A;Title: Comparison of the nucleotide sequence of the SH gene and flanking regions of
A;Reference number: A38479
A;Accession: A38479
A;Molecule type: mRNA
A;Residues: 1-57 <TUR>
R;Takeuchi, K.
submitted to JIPID, November 1990
A;Reference number: JU0308
A;Accession: JU0308
A;Molecule type: mRNA
A;Residues: 1-57 <TAK>
R;Afzal, M.A.; Pickford, A.R.; Forsey, T.; Heath, A.B.; Minor, P.D.
J. Gen. Virol. 74, 917-920, 1993
A;Title: The Jeryl Lynn vaccine strain of mumps virus is a mixture of two distinct
A;Reference number: JQ2366; MUID:93260409; PMID:8492099
A;Accession: JQ2366
A;Molecule type: mRNA
A;Residues: 1-57 <AFZ>
A;Experimental source: strain Urabe

Alignment Scores:
Pred. No.: 110 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x SHNZMV (1-57)

QY 222 TTCCTGCTCCTGATCTCTT 202

Db 12 PheLeuLeuLeuLeuLeu 18
|||||

RESULT 12

SHNZME

small hydrophobic protein - mumps virus (strain Enders)

C;Species: mumps virus
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jun-2000
C;Accession: JU0305; S24824
R;Takeuchi, K.
submitted to JIPID, November 1990
A;Reference number: JU0304
A;Accession: JU0305
A;Molecule type: mRNA
A;Residues: 1-57 <TAK>
A;Cross-references: GB:D90231; NID:g222156; PIDN:BAAL4279.1; PID:g222157
A;Experimental source: strain Enders
R;Yeo, R.P.; Afzal, M.A.; Forsey, T.; Rima, B.K.
submitted to the EMBL Data Library, January 1992
A;Description: Nucleotide sequence analysis of the SH gene of mumps virus reveals se

Alignment Scores:
Pred. No.: 110 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x SHNZME (1-57)

QY 222 TTCCTGCTCCTGATCTCTT 202

|||||

Db 12 PheLeuLeuLeuLeuLeu 18

RESULT 13

SHNZMU

small hydrophobic protein - mumps virus (strain Urabe Am9)

C;Species: mumps virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 07-May-1999
C;Accession: A38479; JU0308; JQ2366; S11637
R;Turner, P.C.; Forsey, T.; Minor, P.D.
J. Gen. Virol. 72, 435-437, 1991
A;Title: Comparison of the nucleotide sequence of the SH gene and flanking regions of
A;Reference number: A38479
A;Accession: A38479
A;Molecule type: mRNA
A;Residues: 1-57 <TUR>
R;Takeuchi, K.
submitted to JIPID, November 1990
A;Reference number: JU0308
A;Accession: JU0308
A;Molecule type: mRNA
A;Residues: 1-57 <TAK>
R;Afzal, M.A.; Pickford, A.R.; Forsey, T.; Heath, A.B.; Minor, P.D.
J. Gen. Virol. 74, 917-920, 1993
A;Title: The Jeryl Lynn vaccine strain of mumps virus is a mixture of two distinct
A;Reference number: JQ2366; MUID:93260409; PMID:8492099
A;Accession: JQ2366
A;Molecule type: mRNA
A;Residues: 1-57 <AFZ>
A;Experimental source: strain Urabe

C:Genetics:

A:Gene: SH
C:Superfamily: mumps virus small hydrophobic protein
C:Keywords: transmembrane protein
F:8-28/Domain: transmembrane #status predicted <TMM>

Alignment Scores:
Pred. No.: 110 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x SHNZMU (1-57)

QY 222 TTCCTGCTCTGTACTTCTT 202

Db 12 PheLeuLeuLeuLeuLeu 18

RESULT 14

SHNZMS

small hydrophobic protein - mumps virus (strain Miyahara)

C:Species: mumps virus

C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jun-2000

C:Accession: JU0304

R:Takeuchi, K.

submitted to JIPID, November 1990

A:Reference number: JU0304

A:Accession: JU0304

A:Molecule type: mRNA

A:Residues: 1-57 <TAK>

A:Cross-references: GB:D90234; NID:g222162; PIDN:BAAL4282.1; PID:g222163

C:Superfamily: mumps virus small hydrophobic protein

C:Keywords: transmembrane protein

F:13-29/Domain: transmembrane #status predicted <TMM>

Alignment Scores:

Pred. No.: 110 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x SHNZMS (1-57)

QY 222 TTCCTGCTCTGTACTTCTT 202

Db 12 PheLeuLeuLeuLeuLeu 18

RESULT 15

SHNZMM

small hydrophobic protein - mumps virus

C:Species: mumps virus

C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jun-2000

C:Accession: JU0307; S19870

R:Takeuchi, K.

submitted to JIPID, November 1990

A:Reference number: JU0304

A:Accession: JU0307

A:Molecule type: mRNA

A:Residues: 1-57 <TAK>

A:Cross-references: GB:D90233; NID:g222160; PIDN:BAAL4281.1; PID:g222161

A:Experimental source: strain Matsuyama

R:Yeo, R.P.; Afzal, M.A.; Forsey, T.; Rima, B.K.

submitted to the EMBL Data Library, January 1992

A:Description: Nucleotide sequence analysis of the SH gene of mumps virus reveals several

A:Reference number: S19866

A:Accession: S19870

A:Molecule type: DNA

A:Residues: 1-6,'L',11-18,'Y',20-27,'VV',30-41,'A',43-47,'LF',50-52,'L',54-57 <Y

A:Cross-references: EMBL:X63713; NID:g60604; PIDN:CAA45246.1; PID:g60605

A:Experimental source: strain Bristoll

C:Genetics:

A:Gene: SH
C:Superfamily: mumps virus small hydrophobic protein
C:Keywords: transmembrane protein
F:8-29/Domain: transmembrane #status predicted <TMM>

Alignment Scores:
Pred. No.: 110 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x SHNZMM (1-57)

QY 222 TTCCTGCTCTGTACTTCTT 202

Db 12 PheLeuLeuLeuLeuLeu 18

Search completed: January 21, 2003, 14:54:45

Job time : 16.7244 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 13:52:06 ; Search time 8.44488 Seconds
(without alignments)
2701.279 Million cell updates/sec

Title: US-10-060-830-6
Perfect score: 90
Sequence: 1 caacttcagtgtgcagcc.....tttgaagatgatgctgcttt 275

Scoring table:
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Word size: 1

Total number of hits satisfying chosen parameters: 225614

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEJ-frame+n2p.model -DEV=xlh
-Q/cgn2_1/USFTO.spool/US10060830/runat_16012003_092701_1487/app_query.fasta_1.1109
-DB-SwissProt_40 -QMT-fastan -SUFFIX-oln2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-oligo -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFTW=pc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10060830 @CGL_1.11 @runat_16012003_092701_1487 -NCPU=3
-NO_XUPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	8	9.0	366	1 YM85_YEAST	Q04806 saccharomyc
C 2	8	8.9	384	1 Y039_MYCPN	P75063 mycoplasma
C 3	7	7.9	57	1 VSH_MUMPI	P19719 mumps virus
C 4	7	7.9	57	1 VSH_MUMPA	P22111 mumps virus
C 5	7	7.9	57	1 VSH_MUMPB	P28082 mumps virus
C 6	7	7.9	57	1 VSH_MUMPE	P22109 mumps virus
C 7	7	7.9	57	1 VSH_MUMPK	P28086 mumps virus
C 8	7	7.9	57	1 VSH_MUMPL	P28083 mumps virus
C 9	7	7.9	57	1 VSH_MUMPM	P22112 mumps virus
C 10	7	7.9	57	1 VSH_MUMPR	P28087 mumps virus
C 11	7	7.9	57	1 VSH_MUMPT	P22113 mumps virus
C 12	7	7.9	57	1 VSH_MUMPU	P20716 mumps virus
C 13	7	7.9	72	1 YVDD_VACCV	P04306 vaccinia vi
C 14	7	7.9	102	1 L08_LYCES	Q43495 lycopersico
C 15	7	7.9	113	1 V62_SPVIR	P15897 spiroplasma
C 16	7	7.8	144	1 MRAZ_STAAU	O07319 staphylococ
C 17	7	7.9	146	1 YN68_YEAST	P42834 saccharomyc
C 18	7	7.8	153	1 YLR3_EBVP3	Q07286 epstein-bar

C 19	7	7.9	172	1 YFIR_ECOLI	P76597 escherichia
C 20	7	7.9	197	1 HS41_PEA	P19244 pisum sativ
C 21	7	7.8	255	1 MYF5_BOVIN	P17667 bos taurus
C 22	7	7.8	255	1 MYF5_HUMAN	P13349 homo sapien
C 23	7	7.8	255	1 MYF5_MOUSE	P24699 mus musculu
C 24	7	7.9	278	1 Y066_CAEEL	P34606 caenorhabd
C 25	7	7.9	314	1 TPIC_FRAAN	Q9m498 fragaria an
C 26	7	7.9	342	1 CGIC_SCHPO	O74627 schizosacch
C 27	7	7.9	380	1 YRP4_SHVX	Q04583 shallot vir
C 28	7	7.9	381	1 CYB_ANTLA	O20434 antelthinomy
C 29	7	7.9	381	1 CYB_DASSP	Q34382 dasyturus sp
C 30	7	7.8	397	1 CBG_MOUSE	Q06770 mus musculu
C 31	7	7.9	417	1 GP61_HUMAN	Q9bzj8 homo sapien
C 32	7	7.9	431	1 OC11_MOUSE	P13362 mus musculu
C 33	7	7.8	435	1 STCB_EMENI	Q12608 emerichia
C 34	7	7.8	438	1 IAPI_DROME	Q24306 drosophila
C 35	7	7.9	477	1 INGR_MOUSE	P15361 mus musculu
C 36	7	7.8	477	1 YBFO_ECOLI	P77779 escherichia
C 37	7	7.9	495	1 NUSA_ECOLI	P03003 escherichia
C 38	7	7.9	500	1 NUSA_SALTY	P37430 salmonella
C 39	7	7.8	518	1 MTCQ_MOUSE	Q60754 mus musculu
C 40	7	7.8	533	1 NCRI_RAT	Q9wub5 rattus norv
C 41	7	7.8	566	1 YGA3_YEAST	P53197 saccharomyc
C 42	7	7.9	673	1 ABG8_HUMAN	Q9h221 homo sapien
C 43	7	7.9	708	1 ICAL_HUMAN	P20810 homo sapien
C 44	7	7.9	784	1 TLR2_MOUSE	Q9qun7 mus musculu
C 45	7	7.8	885	1 CHS3_EXODE	P30502 exophiala d

ALIGNMENTS

RESULT 1
ID YM85_YEAST STANDARD; PRT; 366 AA.
AC Q04806;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 42.4 kDa protein in FAA4-HOR7 intergenic region.
GN YMR251W OR YMR920.05
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE YGR154C/YKR076W/YMR251W (YEAST) /
CC YQJG (E. COLI) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48639; CAA88578.1; -
CC SGD; S0004863; YMR251W.
CC InterPro: IPR004046; GST_Cterm.
CC Hypothetical protein.
SQ SEQUENCE 366 AA; 42403 MW; C0BF9F2D3AB5C15B CRC64;

Alignment Scores:
Pred. No.: 4.79 Length: 366
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.99% Indels: 0
DB: 1 Gaps: 0

; NAME/KEY: SITE
; LOCATION: (1397)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-18

Query Match
Best Local Similarity 5.8%; Score 16; DB 4; Length 1422;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 TTGAAGATGATGCTG 271
Db 977 TTGAAGATGATGCTG 992
|||||

RESULT 14

US-09-484-970B-152/c
; Sequence 152, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 152
; LENGTH: 2030
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 400799.2CB1
; NAME/KEY: unsure
; LOCATION: 294-318, 334, 487, 1603-1922
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-152

Query Match
Best Local Similarity 5.8%; Score 16; DB 4; Length 2030;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 GCTGGGAAGCCAGGTC 156
Db 1114 GCTGGGAAGCCAGGTC 1099
|||||

RESULT 15

US-09-150-460B-5/c
; Sequence 5, Application US/09150460B
; Patent No. 6130882
; GENERAL INFORMATION:
; APPLICANT: Lee, Cheng-Chi
; APPLICANT: Albrecht, Urs
; APPLICANT: Eichele, Gregor
; APPLICANT: Sun, Zhong Sheng
; TITLE OF INVENTION: Mammalian Circadian Rhythm-Like Gene
; FILE REFERENCE: D6039
; CURRENT APPLICATION NUMBER: US/09/150,460B
; CURRENT FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: US 60/058,256
; PRIOR FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 5
; LENGTH: 3057
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence of 3.0 kb cDNA of human RIGUI
US-09-150-460B-5

Query Match
Best Local Similarity 5.8%; Score 16; DB 4; Length 3057;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 ACCAACCTCCCCCACT 64
Db 2924 ACCAACCTCCCCCACT 2909
|||||

Search completed: January 21, 2003, 15:00:45
Job time : 986.079 secs

US-10-060-830-6 (1-275) x YM85_YEAST (1-366)

QY 56 AGGTTGGTCCCGCTGACCTTGAA 33
|||||
Db 194 ArgLeuValProArgSerLeuGlu 201

RESULT 2

Y039_MYCPN STANDARD; PRT; 384 AA.
AC P75063;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG039 homolog (D09_orf384).
GN MFN051 OR MF103.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;

RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Harrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000012; AAB95751.1;
DR InterPro; IPR000927; D_aa_Oxidase.
DR Pfam; PF01266; DAO; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 384 AA; 42724 MW; DA3E128719EE1BCD CRC64;

Alignment Scores:
Pred. No.: 4.77 Length: 384
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.89% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x Y039_MYCPN (1-384)

QY 88 TCTCAGGACTGACGCTGCTCT 111
|||||
Db 348 SerProGlyLeuThrAlaAlaPro 355

RESULT 3

VSH_MUMPI STANDARD; PRT; 57 AA.
AC P19719;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Small hydrophobic protein.
GN SH.
OS Mumps virus (strain SBL-1), and
OS Mumps virus (strain SBL).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX NCBI_TaxID=11173, 33729;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=SBL-1;
RX MEDLINE=89125739; PubMed=2915385;
RT Elango N., Koevamees J., Varsanyi T.M., Norrby E.;
RA "mRNA sequence and deduced amino acid sequence of the mumps virus small hydrophobic protein gene";
RL J. Virol. 63:1413-1415(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SBL;
RX MEDLINE=89243815; PubMed=2718625;
RA Elliott G.D., Afzal M.A., Martin S.J., Rima B.K.;
RT "Nucleotide sequence of the matrix, fusion and putative SH protein genes of mumps virus and their deduced amino acid sequences.";
RL Virus Res. 12:61-75(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SBL;
RX MEDLINE=93168036; PubMed=8435047;
RA Yeo R.P., Afzal M.A., Forsey T., Rima B.K.;
RT "Identification of a new mumps virus lineage by nucleotide sequence analysis of the SH gene of ten different strains.";
RL Arch. Virol. 128:371-377(1993).
RN [4]
RP SEQUENCE FROM N.A.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M25421; AAA50290.1;
DR EMBL; D00663; BAA00562.1;
DR EMBL; X63704; CAA45233.1;
DR PIR; A31884; SHNZMV.
DR PIR; S24823; S24823.
DR PIR; C60004; C60004.
DR InterPro; IPR001477; SH.
DR Pfam; PF01445; SH; 1.
DR ProDom; PD001504; SH; 1.
KW Transmembrane.
FT TRANSMEM 8 32 POTENTIAL.
SQ SEQUENCE 57 AA; 6718 MW; D344CB495BBEA323 CRC64;

Alignment Scores:
Pred. No.: 59.1 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x VSH_MUMPI (1-57)

QY 222 TTCTGCTCCTGATCTCTT 202
|||||
Db 12 PheLeuLeuLeuLeuLeuLeu 18

RESULT 4

VSH_MUMPI STANDARD; PRT; 57 AA.
AC P22111;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Small hydrophobic protein.
GN SH.
OS Mumps virus (strain Matsuyama).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX NCBI_TaxID=11165;
RN [1]
RP SEQUENCE FROM N.A.

```
RA Takeuchi K.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D50233; BAA14281.1; -
DR PIR; JU0307; SHNZME.
DR InterPro; IPR001477; SH.
DR Pfam; PF01445; SH; 1.
DR ProDom; PD001504; SH; 1.
DR Transmembrane.
KW TRANSMEM
FT TRANSMEM 8 32 POTENTIAL.
SQ SEQUENCE 57 AA; 6697 MW; 1D994E57AA55ECF0 CRC64;
Alignment Scores:
Pred. No.: 59.1 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0
US-10-060-830-6 (1-275) x VSH_MUMPA (1-57)
QY 222 TTCCTGCTCTGATCTTCTT 202
Db 12 PheLeuLeuLeuLeuLeuLeu 18
RESULT 5
VSH_MUMPB
ID VSH_MUMPB - STANDARD; PRT; 57 AA.
AC P28082;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Small hydrophobic protein.
GN SH.
OS Mumps virus (strain Belfast).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses;
OX NCBI_TaxID=11166;
RN [1]
RP SEQUENCE FROM N.A.
RA Yeo R.P., Afzal M.A., Forsey T., Rima B.K.;
RT "Identification of a new mumps virus lineage by nucleotide sequence
RT analysis of the SH gene of ten different strains.";
RL Arch. Virol. 128:371-377(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63709; CAA45242.1; -
DR PIR; S19866; SHNZBF.
DR InterPro; IPR001477; SH.
DR Pfam; PF01445; SH; 1.
DR ProDom; PD001504; SH; 1.
DR Transmembrane.
KW TRANSMEM
FT TRANSMEM 8 32 POTENTIAL.
SQ SEQUENCE 57 AA; 6814 MW; B8F2F917B4A8EF3C CRC64;
Alignment Scores:
```

```
Pred. No.: 59.1 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0
US-10-060-830-6 (1-275) x VSH_MUMPB (1-57)
QY 222 TTCCTGCTCTGATCTTCTT 202
Db 12 PheLeuLeuLeuLeuLeuLeu 18
RESULT 6
VSH_MUMPE
ID VSH_MUMPE STANDARD; PRT; 57 AA.
AC P22109;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Small hydrophobic protein.
GN SH.
OS Mumps virus (strain Enders).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX NCBI_TaxID=11167;
RN [1]
RP SEQUENCE FROM N.A.
RA Takeuchi K.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93168036; PubMed=8435047;
RA Yeo R.P., Afzal M.A., Forsey T., Rima B.K.;
RT "Identification of a new mumps virus lineage by nucleotide sequence
RT analysis of the SH gene of ten different strains.";
RL Arch. Virol. 128:371-377(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90231; BAA14279.1; -
DR EMBL; X63705; CAA45235.1; -
DR PIR; JU0305; SHNZME.
DR PIR; S24824; S24824.
DR InterPro; IPR001477; SH.
DR Pfam; PF01445; SH; 1.
DR ProDom; PD001504; SH; 1.
DR Transmembrane.
KW TRANSMEM 8 32 POTENTIAL.
SQ SEQUENCE 57 AA; 6717 MW; D443CB495BBBF323 CRC64;
Alignment Scores:
Pred. No.: 59.1 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0
US-10-060-830-6 (1-275) x VSH_MUMPE (1-57)
QY 222 TTCCTGCTCTGATCTTCTT 202
Db 12 PheLeuLeuLeuLeuLeuLeu 18
RESULT 7
VSH_MUMPK
```

```
ID VSH_MUMPK STANDARD; PRT; 57 AA.
AC P28086;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Small hydrophobic protein.
GN SH.
OS Mumps virus (strain Kilham).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
OX NCBI_TaxID=111179;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93168036; PubMed=8435047;
RA Yeo R.P., Afzal M.A., Forsey T., Rima B.K.;
RT "Identification of a new mumps virus lineage by nucleotide sequence
RT analysis of the SH gene of ten different strains.";
RL Arch. Virol. 128:371-377(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63706; CAA45237.1;
DR PIR; S24825; S24825.
DR InterPro; IPR001477; SH.
DR Pfam; PF01445; SH; 1.
DR ProDom; PD001504; SH; 1.
DR ProDom; PD001504; SH; 1.
KW Transmembrane.
FT TRANSMEM 8 32
SQ SEQUENCE 57 AA; 6804 MW; 9BE7204940BA433F CRC64;
POTENTIAL.

Alignment Scores:
Pred. No.: 59.1 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x VSH_MUMPK (1-57)

QY 222 TTCCTGCTCCTGATCTTCT 202
|||||
Db 12 PheLeuLeuLeuLeuLeu 18

RESULT 8
VSH_MUMPL
ID VSH_MUMPL STANDARD; PRT; 57 AA.
AC P28083;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Small hydrophobic protein.
GN SH.
OS Mumps virus (strain Bristol 1).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
OX NCBI_TaxID=111170;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93168036; PubMed=8435047;
RA Yeo R.P., Afzal M.A., Forsey T., Rima B.K.;
RT "Identification of a new mumps virus lineage by nucleotide sequence
RT analysis of the SH gene of ten different strains.";
RL Arch. Virol. 128:371-377(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63713; CAA45246.1;
DR PIR; S19870; SHNZBL.
DR InterPro; IPR001477; SH.
DR Pfam; PF01445; SH; 1.
DR ProDom; PD001504; SH; 1.
DR ProDom; PD001504; SH; 1.
KW Transmembrane.
FT TRANSMEM 8 32
SQ SEQUENCE 57 AA; 6753 MW; 8EF2EC57BA4FBC1D CRC64;
POTENTIAL.

Alignment Scores:
Pred. No.: 59.1 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x VSH_MUMPL (1-57)

QY 222 TTCCTGCTCCTGATCTTCT 202
|||||
Db 12 PheLeuLeuLeuLeuLeu 18

RESULT 9
VSH_MUMPM
ID VSH_MUMPM STANDARD; PRT; 57 AA.
AC P22112;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Small hydrophobic protein.
GN SH.
OS Mumps virus (strain Miyahara vaccine).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
OX NCBI_TaxID=111171;
RN [1]
RP SEQUENCE FROM N.A.
RA Takeuchi K.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90234; BAA14282.1;
DR PIR; J00304; SHNZMS.
DR InterPro; IPR001477; SH.
DR Pfam; PF01445; SH; 1.
DR ProDom; PD001504; SH; 1.
DR ProDom; PD001504; SH; 1.
KW Transmembrane.
FT TRANSMEM 8 32
SQ SEQUENCE 57 AA; 6621 MW; BC794E57A4A0F9E6 CRC64;
POTENTIAL.

Alignment Scores:
Pred. No.: 59.1 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x VSH_MUMPM (1-57)
```

```
QY 222 TTCCTGCTCCTGATCTTCTT 202
Db 12 PheLeuLeuLeuLeuLeu 18

RESULT 10
VSH_MUMPR
ID VSH_MUMPR STANDARD; PRT; 57 AA.
AC P28087;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Small hydrophobic protein.
GN SH.
OS Mumps virus (strain RW).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
OX NCBI_TaxID=11172;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93168036; PubMed=8435047;
RA Yeo R.P., Afzal M.A., Forsey T., Rima B.K.;
RT "Identification of a new mumps virus lineage by nucleotide sequence
RT analysis of the SH gene of ten different strains.";
RL Arch. Virol. 128:371-377(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X63708; CAA45241.1;
CC PIR; S24827; S24827; SH.
CC InterPro; IPR001477; SH.
CC Pfam; PF01445; SH; 1.
CC ProDom; PD001504; SH; 1.
CC Transmembrane.
CC TRANSMEM 8 32 POTENTIAL.
CC SEQUENCE 57 AA; 6892 MW; B9FA5B0945BC04D3 CRC64;

Alignment Scores:
Pred. No.: 59.1 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x VSH_MUMPR (1-57)

QY 222 TTCCTGCTCCTGATCTTCTT 202
Db 12 PheLeuLeuLeuLeuLeu 18

RESULT 11
VSH_MUMPT
ID VSH_MUMPT STANDARD; PRT; 57 AA.
AC P22113;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Small hydrophobic protein.
GN SH.
OS Mumps virus (strain Takahashi).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
OX NCBI_TaxID=11174;
RN [1]
RP SEQUENCE FROM N.A.
RX Takeuchi K.;
```

```
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; D90235; BAA14283.1;
CC PIR; JU0309; SHN2MT.
CC InterPro; IPR001477; SH.
CC Pfam; PF01445; SH; 1.
CC ProDom; PD001504; SH; 1.
CC Transmembrane.
CC TRANSMEM 8 32 POTENTIAL.
CC SEQUENCE 57 AA; 6745 MW; EC965E46AB45ECF0 CRC64;

Alignment Scores:
Pred. No.: 59.1 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x VSH_MUMPT (1-57)

QY 222 TTCCTGCTCCTGATCTTCTT 202
Db 12 PheLeuLeuLeuLeuLeu 18

RESULT 12
VSH_MUMPU
ID VSH_MUMPU STANDARD; PRT; 57 AA.
AC P20716;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE Small hydrophobic protein.
GN SH.
OS Mumps virus (strain Urabe vaccine AM9).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
OX NCBI_TaxID=11175;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91132144; PubMed=1993881;
RA Turner P.C., Forsey T., Minor P.D.;
RT "Comparison of the nucleotide sequence of the SH gene and flanking
RT regions of mumps vaccine virus (Urabe strain) grown on different
RT substrates and isolated from vaccinees.";
RL J. Gen. Virol. 72:435-437(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Takeuchi K.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X52816; CAA37000.1;
CC PIR; D90236; BAA14284.1;
CC InterPro; IPR001477; SH.
CC Pfam; PF01445; SH; 1.
CC ProDom; PD001504; SH; 1.
```

```
KW Transmembrane. 8 32 POTENTIAL.
SQ SEQUENCE 57 AA; 6755 MW; AC784E440AD92CE6 CRC64;

Alignment Scores:
Pred. No.: 59.1 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x VSH_MUMPU (1-57)

QY 222 TTCCTGCTCTGATCTTCTT 202
Db 12 PheLeuLeuLeuLeuLeu 18

RESULT 13
YVDD_VACCV STANDARD; PRT; 72 AA.
AC P04306;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 7.9 kDa protein.
GN D ORF D.
OS Vaccinia virus (strain WR), and
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10254, 10249;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WR;
RX MEDLINE=86291159; PubMed=3739227;
RA Niles E.G., Condit R.C., Caro P., Davidson K., Matusick L., Seto J.;
RT "Nucleotide sequence and genetic map of the 16-kb vaccinia virus
RT HindIII D fragment.";
RL Virology 153:96-112(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Copenhagen;
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [3]
RP COMPLETE GENOME.
RC STRAIN=Copenhagen;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; M15058; AAA48260.1; -
DR EMBL; M35027; AAA48103.1; -
DR PIR; A03880; QQV29.
DR PIR; G42516; G42516.
KW Hypothetical protein.
SQ SEQUENCE 72 AA; 7877 MW; 40094F7C5DEA9AC1 CRC64;

Alignment Scores:
Pred. No.: 58.1 Length: 72
```

```
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x YVDD_VACCV (1-72)

QY 222 TTCCTGCTCTGATCTTCTT 202
Db 6 PheLeuLeuLeuLeuLeu 12

RESULT 14
108_LYCES STANDARD; PRT; 102 AA.
AC Q43495;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Protein 108 precursor.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. VF36; TISSUE=Anther;
RX MEDLINE=94143497; PubMed=8310077;
RA Chen R., Smith A.G.;
RT "Nucleotide sequence of a stamen- and tapetum-specific gene from
RT Lycopersicon esculentum.";
RL Plant Physiol. 101:1413-1413(1993).
CC -1- TISSUE SPECIFICITY: STAMEN- AND TAPETUM-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO THE A9 / FILL FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; Z14088; CAA78466.1; -
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Try/amyl_inhbt.
DR Pfam; PF00234; tryp_alpha_amyl; 1.
DR SMART; SM00499; AAI; 1.
KW Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 102 PROTEIN 108.
FT DISULFID 41 77 BY SIMILARITY.
FT DISULFID 51 66 BY SIMILARITY.
FT DISULFID 67 92 BY SIMILARITY.
FT DISULFID 79 99 BY SIMILARITY.
SQ SEQUENCE 102 AA; 10576 MW; CFBAA1231C3A5E92 CRC64;

Alignment Scores:
Pred. No.: 56.5 Length: 102
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x 108_LYCES (1-102)

QY 219 CTGCTCTGATCTTCTTGTG 199
Db 19 LeuLeuLeuLeuLeuVal 25

RESULT 15
```

```

VG6_SPV1R
ID VG6_SPV1R STANDARD; PRT; 113 AA.
AC P15897;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Gene 6 protein.
GN 6.
OS Spiroplasma virus SpV1-R8A2 B.
OC Viruses; ssDNA viruses; Inoviridae; Plectrovirus.
OX NCBI_TaxID=10854;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206799; PubMed=2320423;
RA Renaudin J., Aulio P., Vignault J.C., Bove J.M.;
RT "Complete nucleotide sequence of the genome of Spiroplasma citri
virus SpV1-R8A2 B.";
RL Nucleic Acids Res. 18:1293-1293(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51344; CAA35733.1;
SQ SEQUENCE 113 AA; 13626 MW; F813E7F4FA80A909 CRC64;

Alignment Scores:
Pred. No.: 56.1 Length: 113
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.87% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-6 (1-275) x VG6_SPV1R (1-113)
QY 225 CCCTTCTGCTCTGATACCTT 205
Db 60 ProPheLeuLeuLeuLeu 66

```

Search completed: January 21, 2003, 14:49:57
Job time : 10.4449 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:43:31 ; Search time 28.3661 Seconds
(without alignments)
3995.112 Million cell updates/sec

Title: US-10-060-830-6
Perfect score: 90
Sequence: 1 caacttcagttggtcagccc.....tttgaagatgatgtgctttt 275

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Word size: 1
Total number of hits satisfying chosen parameters: 1343044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10060830/runat.16012003_092702.1498/app_query.fasta_1.1109
-DB=SPTREMBL_21 -OPMT=fastan -SUFFIX=olin2p.rspt -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdd
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830@cgn.1.1.57_@runat.16012003_092702.1498 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SPTREMBL_21:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phage:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_rviro:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	85	94.4	775	4	Q96PD2 homo sapien

2	85	94.4	775	4	Q8TDX2
3	17	18.9	251	11	Q9D9K5
4	17	18.9	769	11	Q912V3
5	17	18.9	769	11	Q912V2
6	8	9.0	187	11	Q99JG4
7	8	9.0	190	16	Q8XJF7
8	8	8.9	210	16	Q8UJ11
9	8	8.9	242	10	Q9FSL1
10	8	8.9	258	16	Q928D8
11	8	9.0	291	10	Q9AWZ0
12	8	8.9	361	17	Q9YFD2
13	8	8.9	362	16	Q9XA18
14	8	8.9	453	17	Q28941
15	8	8.9	477	16	Q8XNC9
16	8	8.9	479	16	Q9XIE6
17	8	9.0	512	16	Q84760
18	8	9.0	671	11	Q99NI0
19	8	9.0	671	11	Q8R499
20	8	8.9	673	2	Q8RNU7
21	8	9.0	685	11	Q99MW6
22	8	8.9	826	16	Q82122
23	8	8.9	971	3	Q9P953
24	8	8.9	1162	5	Q9VB40
25	8	8.9	5157	3	Q01135
26	7	7.9	15	4	Q9UMV6
27	7	7.8	44	6	Q02774
28	7	7.8	54	11	Q91VR1
29	7	7.9	57	12	Q89538
30	7	7.9	57	12	Q83733
31	7	7.9	57	12	Q89593
32	7	7.9	57	12	Q89564
33	7	7.9	57	12	Q88348
34	7	7.9	57	12	P89006
35	7	7.9	57	12	P89007
36	7	7.9	57	12	P89009
37	7	7.9	57	12	P89011
38	7	7.9	57	12	P89014
39	7	7.9	57	12	P89015
40	7	7.9	57	12	P89016
41	7	7.9	57	12	P89018
42	7	7.9	57	12	P89019
43	7	7.9	57	12	O41006
44	7	7.9	57	12	Q9WAF2
45	7	7.9	57	12	Q9WAF7

ALIGNMENTS

RESULT 1					
Q96PD2	ID	Q96PD2	PRELIMINARY;	PRT;	775 AA.
AC	Q96PD2;				
DT	01-DEC-2001	(TREMBlrel. 19, Created)			
DT	01-DEC-2001	(TREMBlrel. 19, Last sequence update)			
DT	01-JUN-2002	(TREMBlrel. 21, Last annotation update)			
DE	Endothelial and smooth muscle cell-derived neuropilin-like protein.				
GN	ESDN.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
ON	NCBI_TaxID=9606;				
RX	[1]				
RP	SEQUENCE FROM N.A.				
RX	PubMed=11447234;				
RA	Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,				
RA	Matsumori A., Sasayama S., Honjo T., Tashiro K.;				
RT	*ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular				
RT	Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is				
RT	Up-regulated after Vascular Injury.;				
RL	J. Biol. Chem. 276:34105-34114(2001).				
CC	- - SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL REGULATORS.				

DR	EMBL; AF397547; AAL30178.1; -;
DR	InterPro; IPR000859; CUB_domain.
DR	InterPro; IPR000421; FA5B_C.
DR	InterPro; IPR004043; LCCL_dom.
DR	Pfam; PF00431; CUB_1.
DR	Pfam; PF00754; F5_F8_type-C; 1.
DR	PROSITE; PS01180; CUB; 1.
KW	DNA-binding; Transcription regulation.
SQ	SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;
 Alignment Scores:	
Pred. No.:	2.54e-79
Score:	85.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	94.44%
DB:	4
 US-10-060-830-6 (1-275) x Q96PD2 (1-775)	
Qy	3 ACTTCAGTTGTCACGCCCTCCACATCCACTTTCACGGCTACGGGAACCAACCTCCCCA 62
Db	691 ThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProProPro 710
Qy	63 CTATGGGAACCTTACAATACACTCTCTCCAGGACTGCACAGCTGTCTCCTCAGCCCGCC 122
Db	711 LeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAla 730
Qy	123 CAGTATCATCCCCGAAGCTGGAGCCAGCTGCTACCTGCCCCAGCACGAATTGGTGAC 182
Db	731 GlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTyr 750
Qy	183 CAGGTGCCACAGACACACAAGAAGTATCAGGAGCAGGAGGATGGGGAAATGTGATGTT 242
Db	751 GlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGlyGluCysaspVal 770
Qy	243 TTTAAGAANAATCCTT 257
Db	771 PheLysGluIleLeu 775
 RESULT 2	
Q8TDX2	ID Q8TDX2 PRELIMINARY; PRT; 775 AA.
AC	Q8TDX2;
DT	01-JUN-2002 (TrEMBLrel. 21, Created)
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	CLCP1.
GN	CLCP1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=LUNG;
RA	Koshikawa K., Osada H., Kozaki K., Konishi H., Masuda A.,
RA	Tatematsu Y., Mitsudomi T., Nakao A., Takahashi T.;
RT	*significant up-regulation of a novel gene, CLCP1, in a highly
RT	metastatic lung cancer subline as well as in lung cancers in vivo.*;
EL	Oncogene 0:0-0(2002).
DR	EMBL; AB073146; BAB91138.1; -;
SQ	SEQUENCE 775 AA; 85073 MW; 6D07223B21BE5A42 CRC64;
 Alignment Scores:	
Pred. No.:	2.54e-79
Score:	85.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	94.44%
DB:	4
 US-10-060-830-6 (1-275) x Q8TDX2 (1-775)	

```
Db 188 LeuValGlyThrTyraAsnThrLeuLeuSerArgThrAspSerCysSerSer 204
RESULT 4
ID Q912V3 PRELIMINARY; PRT; 769 AA.
AC Q912V3;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like
DE protein.
GN ESDN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR;
RX PubMed=11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury.";
RL J. Biol. Chem. 276:34105-34114(2001).
CC -/- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AF387548; AAL301179.1; -.
DR MGD; MGI:1920629; Esdn.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
KW DNA-binding; Transcription regulation.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;
US-10-060-830-6 (1-275) x Q912V3 (1-769)

Alignment Scores:
Pred. No.: 8.13e-09 Length: 769
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 18.89% Indels: 0
DB: 11 Gaps: 0

US-10-060-830-6 (1-275) x Q912V3 (1-769)

QY 63 CTAGTGGGAACCTACAACTCTCTCCAGGACTGACAGCTGCTCCTCA 113
|||||
Db 706 LeuValGlyThrTyraAsnThrLeuLeuSerArgThrAspSerCysSerSer 722
RESULT 5
ID Q912V2 PRELIMINARY; PRT; 769 AA.
AC Q912V2;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like
DE protein.
GN ESDN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX PubMed=11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
```

```
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury.";
RL J. Biol. Chem. 276:34105-34114(2001).
CC -/- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AF387549; AAL30180.1; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
KW DNA-binding; Transcription regulation.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 769 AA; 83915 MW; 5B7936C8FA063A93 CRC64;
US-10-060-830-6 (1-275) x Q912V2 (1-769)

Alignment Scores:
Pred. No.: 8.13e-09 Length: 769
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 18.89% Indels: 0
DB: 11 Gaps: 0

US-10-060-830-6 (1-275) x Q912V2 (1-769)

QY 63 CTAGTGGGAACCTACAACTCTCTCCAGGACTGACAGCTGCTCCTCA 113
|||||
Db 706 LeuValGlyThrTyraAsnThrLeuLeuSerArgThrAspSerCysSerSer 722
RESULT 6
ID Q99JG4 PRELIMINARY; PRT; 187 AA.
AC Q99JG4;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Nuclear export factor-b (Fragment).
DE Nuclear export factor-b (Fragment).
GN NXF2 OR NXF-B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS; TISSUE=BRAIN;
RC Froyen G.F.V.;
RA Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS; TISSUE=BRAIN;
RA Lin J.;
RT "NXF5, a novel member of the nuclear RNA export factor family,
RT Involved in mental retardation.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ305319; CAC36304.1; -.
DR MGD; MGI:1933192; Nxf2.
FT NON_TER 1
SQ SEQUENCE 187 AA; 21548 MW; ACFC03D4FCA9FD24 CRC64;
US-10-060-830-6 (1-275) x Q99JG4 (1-187)

Alignment Scores:
Pred. No.: 21.3 Length: 187
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.99% Indels: 0
DB: 11 Gaps: 0

US-10-060-830-6 (1-275) x Q99JG4 (1-187)

QY 226 TCCTTCCTGCTGCTGCTGCTGCTTCT 203
|||||
```

Db 120 SerLeuProAlaProAspThrSer 127

RESULT 7

Q8XJF7 PRELIMINARY; PRT; 190 AA.

AC Q8XJF7

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Hypothetical protein CPE1802.

GN CPE1802.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Clostridiaceae; Clostridium.

OX NCBI_TaxID=1502;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=13 / TYPE A;

RX PubMed=11792842;

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yanashita A.,

RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;

RT "Complete genome sequence of Clostridium perfringens, an anaerobic

RT flesh-eater."

RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

DR EMBL: AP003191; BAB81508.1; -

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 190 AA; 22339 MW; B70C34F8EAFD3BCF CRC64;

Alignment Scores:

Pred. No.:	21.3	Length:	190
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.99%	Indels:	0
DB:	16	Gaps:	0

US-10-060-830-6 (1-275) x Q8XJF7 (1-190)

Qy 222 TTCCTGCTCCTGACTTGTG 199

Db 8 PheLeuLeuLeuLeuVal 15

RESULT 8

Q8UJ11 PRELIMINARY; PRT; 210 AA.

AC Q8UJ11

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Transcriptional regulator, TetR family.

GN ATU5496 OR AGR-PAT_729.

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OG Plasmid AT.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Rhizobium.

OX NCBI_TaxID=176299;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21608550; PubMed=11743193;

RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,

RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,

RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,

RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,

RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,

RA Raymond C., Rouse G., Saenphumachak C., Wu Z., Romero P., Gordon D.,

RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,

RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,

RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,

RA Nester E.W.;

RT "The genome of the natural genetic engineer Agrobacterium tumefaciens

RT C58."

RL Science 294:2317-2323(2001).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21608551; PubMed=11743194;

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,

RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,

RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,

RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,

RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,

RA Cielo C., Slater S.;

RT "Genome sequence of the plant pathogen and biotechnology agent

RT Agrobacterium tumefaciens C58."

RL Science 294:2323-2328(2001).

DR EMBL: AF008970; AAL46182.1; ALT_INIT.

DR EMBL: AE007918; AAK90871.1; -

KW Plasmid; Complete proteome.

SO SEQUENCE 210 AA; 23447 MW; 449934BD70D0F9CF CRC64;

Alignment Scores:

Pred. No.:	21	Length:	210
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.89%	Indels:	0
DB:	16	Gaps:	0

US-10-060-830-6 (1-275) x Q8UJ11 (1-210)

Qy 2 AACTTCAGTGTGTCAGCCCTCCAC 25

Db 9 AsnPheSerTrpSerAlaLeuHis 16

RESULT 9

Q9FSL1 PRELIMINARY; PRT; 242 AA.

AC Q9FSL1

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE Hypothetical 26.7 kDa protein.

GN H0806H05.9.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RA Han B., Feng Q., Huang Y.C., Chen Z.H., Zhou B., Li Y., Zhu J.J.,

RA Tang Y.S., Zhao Q., Liu Y.L., Mu J., Yu Z., Fan D.L., Chen L.,

RA Wang Q.J., Zhang L., Lu Y.Q., Yu S.L., Zhu J., Liu X.H., Hu X.,

RA Lei H.Y., Zhang Y.J., Wang R., Li C., Lu Y., Chen X.C., Zhang Y.,

RA Hu H., Jia P.X., Li T., Qian Y.M., Ying K., Hong G.F.;

RT "Oryza sativa indica (Guangluai4) genomic DNA, chromosome 4, BAC

RT clone: H0806H05."

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL442113; CAC09476.1; -

KW Hypothetical protein.

SO SEQUENCE 242 AA; 26661 MW; F76BF1E959FDACF5 CRC64;

Alignment Scores:

Pred. No.:	20.5	Length:	242
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	8.89%	Indels:	0
DB:	10	Gaps:	0

US-10-060-830-6 (1-275) x Q9FSL1 (1-242)

Qy 100 ACAGCTGCTCTCAGCCAGGCC 123

Db 204 ThrAlaAlaProGlnProArgPro 211

RESULT 10

QY	76	ATGCTCTGTTCTCTCTG-----CNCtTACTGTCTCTGCTCCTGCTGCTCGAGACGCT	129
DB	1	MetaspHethPheProLeuThrTpValPheLeuAlaLeuTyrPheSerArgHisGlnVal	20
QY	130	GGAGCCCAAGGTGATGGATGGACACACTGACTAGTGGCCCTGAGAGTGGACACCTT	189
DB	21	ArgGlyGlnProAspProCysGlyGlyArgLeuAsnSerLysAspAlaGlyTyrIle	40
QY	190	ACATCCATAACTACCCACAGACCTATCCACACAGCACTGTTTGTGAATGGAGATCCGT	249
DB	41	ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluTrpIleValTyr	60
QY	250	GTA---AAGATGGAGAGAGAGTTCATCAAAATTGGT---GACTTTTGACATTGAAGAT	303
DB	61	AlaProGluProAsnGlnIlysIleValLeuAsnPheAsnProHisPheGluIleGluLys	80
QY	304	TCTGATCTTTGTCACCTTAATTAATTACTTTGAGAATTTATAATGGAAATGGAGTCACT	363
DB	81	HisAsp---CysLysTyrAspPheIleGluIleArgAspGlyAspSerGluSerAlaAsp	99
QY	364	GAATAGGCAAAATCACTGGCTCGGGTTCGAAATGAAC-----CATTCAATTGAA	414
DB	100	LeuLeuGlyLysHisGly-----AsnIleAlaProProThrIleIle	114
QY	415	TCAAAGGCAATCAAAATCACTGCTGTTTCATGAGTGAATCATGTTTCTGGAGCGGA	474
DB	115	SerSerGlySerMetLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly	134
QY	475	TTTTTGGCCCTCACTCTGTTATAGATAACAAGATCTAATTACTGTTTGGACACTGCA	534
DB	135	PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer	150
QY	535	TCCAATTTTTTGGAACTGAGTTCAGTACTGCCAGCTGCCGCTGTGCTGTCCTTTT	594
DB	151	LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe	165
QY	595	GCTGAGATATCTGGAACAATTCCTCATGATATAGAGATTCCTCGCCATTCGTGATGGCT	654
DB	166	ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla	181
QY	655	GGTGTGCATCCAGGATAGTGTCAACACAGTTGGCGGCCCAATCAGT-----	702
DB	182	LysProLysMetGluIleIleLeuGlnPheLeuIlePheAspLeuGluHisAspProLeu	201
QY	703	-----GTTGTAATTAGTAAGGATATCC-----	726
DB	202	GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis	221
QY	727	-----TATTATGAAGTCTTTGGTAAACAAGTCACATCTGTG	765
DB	222	ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerGluLeuArgSerSer	241
QY	766	GTGGACACTTATCTACAGTCTTTTACA-----	795
DB	242	ThrGlyIleLeuSerLeuThrPheHisThrAspMetAlaValAlaLysAspGlyPheSer	261
QY	796	-----TTTAAGACAAGTGGATGTTTGGACACTG	825
DB	262	AlaArgTyrTyrLeuValHisGlnGluProLeuGlnAsnPheGlnCysAsnValProLeu	281
QY	826	GGATGGAGTCTGGTGTGATCGCGGATCTCAATTAACAGCATCATCTGCTGGAGTGG	885
DB	282	GlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer-----	297
QY	886	ACTGACCACACAGGCAAGAGAACAGTTGGAAACCCAAAAGCCAGCGCTGAAAAACCT	945
DB	298	-----ThrTyrSerAspGlyArgTyrThrProGlnGlnSerArgLeuHisGlyAsp	314
QY	946	GGACCGCTTGGGCTGCTTTGGCACTGATGAATACCACTGGTTACAAATAGATTGNAAT	1005
DB	315	AspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnValIleLysLeuArg	334

QY	1006	RAGGAAAGAAAATACAGGCATTATAACCACCTGGA-----TCCACCATTGGTGGAGCAC	1059
Db	335	PheLeuThrMetLeuThrAlaLeuThrAlaLeuThrGlnGlyAlaIleSerArgGluThrGlnAsn	354
QY	1060	AATTACTATGTCTGCTCCACAGAAATCCTGTACAGATGATGGCAGCAAAATGGACTGTG	1119
Db	355	GlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAspTrpMetVal	374
QY	1120	TACAGAGAGCCGTGGTGGAGCAAGATAGATATATTTCAAGGAAACAAAGATATTCACCA	1179
Db	375	TyrArgHisGly-----LysAsnHisLysValPheGlnAlaAsnAspAlaThrGlu	392
QY	1180	GATGTGCGTAATAACTTTTGGCCACCAATATTGCACGTTTATTATTAGATGAATCCTACC	1239
Db	393	ValValLeuAsnLysLeuHisAlaProLeuLeuThrArgPheValArgIleArgProGln	412
QY	1240	CAATGGCAGCAGAAAATGCCATGAAATGGAGCTGCTCGGATGTCAAGTTTATTCTCTAAA	1299
Db	413	ThrTrpHisSerGlyIleAlaLeuArgLeuGluLeuPheGlyCys-----	427
QY	1300	GGTCGTCTCCAAAATTACTCAACCTCCACCTCTCGGAACAGCAATGACCTCAAAAAC	1359
Db	428	-----ArgValThrAspAlaProCysSerAsnMetLeuGlyMetLeuSerGly	443
QY	1360	ACTACAGCCCTCCAAAATACCCAAAGTCGTGCCCCCAAAATTT-----	1404
Db	444	LeuIleAlaAspSerGlnIleSerAlaSerSerThrGlnGluTyrLeuTrpSerProSer	463
QY	1405	ACGCAACCACTACAACCTCGCAGTAGCAATGAATTTCTTCGACAGACAGAACAAACAAC	1464
Db	464	AlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgIleProGlnAlaGlnPro	483
QY	1465	GCCAGTCTGTATACAGAAATACTACCGTAACCTCCAAATGTATACCAAGATGTA-----	1518
Db	484	GlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGlyValIleIle	503
QY	1519	-----CGCTGGCTGCAGTCTTCTTGTCCCTGTCTGCTCATCGTCTGCTG	1557
Db	504	GlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPheValaArgLys	523
QY	1558	CTCACTACTCTCATCTCATATTAGTGTGCTGTGGCACTGG-----AGA	1602
Db	524	PheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAspProArgThr	543
QY	1603	AACAGAAAGAAAACACTGAAGCACC-----TATGACTTACCTTACTGGGACCGGCA	1656
Db	544	GlnGlnProLysLeuPheGluGlyAsnMetHisTyrAspThrPro-----	558
QY	1657	GTTGGTGGAAAGGAATGACAGCAATTCCTCTCGAAAAGCAGTGGACCATGAGAAACC	1716
Db	559	-----AspIleArgArgPheAspPro-----Ile	566
QY	1717	CCAGTTGCTATAGCAGCAGCGAAGTTAATCACTCAGTCCAGAGAGATCACCACAGTG	1776
Db	567	ProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIleGlyMetArg	586
QY	1777	CTCAGGCTGACTCTGCAGAGTAT-----GCTCAGCCACTGTTAGGAGAAATGTTGGT	1830
Db	587	LeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGlnThrLeuGlyPro	606
QY	1831	ACATTCATCAAGATCTACCTTTAAACCA-----GAAGAAGAAAGAACGA	1878
Db	607	ThrValLysSerGluGluThrThrProTyrProThrGluGluAlaThrGluCys	626
QY	1879	GGC-----TATCCAGCTATAGATCTTACAATCACTCAGGCGGACGAGGAG--	1921
Db	627	GlyGluAsnCysSerPheGluAspLysAsp-LeuGlnLeuProSerGlyPheAsnCys	646
QY	1922	-----TTTATCATGCCT-----ATGCTGAACCATCCCAA---TTACG	1956
Db	646	AsnPheAspPheLeuGluGluProCysGlyTrpMetTyrAspHisAlaLysTrpLeuArg	666
QY	1957	GGCGCTAGTATGCAACCCCAATCATCATGGACATGTCAGGCGACCCCAACAA-----	2008

```

Db 299 ThrAsnTrpSerAlaGluArgSerArgLeuAsnTyProGluAsnGlyTrpThrProGly 318
Qy 967 GCCACTGATGATACCAAGTGGTTACAAATAGATATTGTAATAGGAAAGAAATACACAGC 1026
Db 319 GluAspSerTyArgGluTrpIleGlnValAspLeuGlyLeuLeuArgPheValThrAla 338
Qy 1027 ATTATAACCACTGGA-----TCCACCATGGTGGAGCAACATTAATGCTGCTGCCTAC 1080
Db 339 ValGlyThrGlnGlyAlaIleSerLysGluThrLysLysLysTyTyTyValLysThrTy 358
Qy 1081 AGAATCCTGTACATGATGATGGCAGAAATGGACTGTGTACAGAGACCGTGGTGGAG 1140
Db 359 LysIleAspValSerSerAsnGlyGluAspTrpIleThrLysGluGly-----Asn 376
Qy 1141 CAAGATAAGATATTCAAGGAAACAAAGATATCACCAGGATGTCGGAATAACATTTTG 1200
Db 377 LysProValLeuPheGlnGlyAsnThrAsnProThrAspValValAlaValPhePro 396
Qy 1201 CCACCAATATTGACAGCTTTTATTAGATGATGATCCTACCCAGCAGACAGAAATGCC 1260
Db 397 LysProLeuIleThrArgPheValArgIleLysProAlaThrTrpGluThrGlyIleSer 416
Qy 1261 ATGAAATGGAGCTGCTCGGATGTCAGTTTATTCCTAAAGTGTCTCCCAAACTTACT 1320
Db 417 MetArgPheGluValTyGlyCys-----LysIleThr 427
Qy 1321 CAACCTCCA 1329
Db 428 AspTyPro 430

RESULT 4
NRP2_HUMAN
ID NRP2_HUMAN STANDARD; PRT; 931 AA.
AC O60462; O14820; O14821;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165
DE receptor 2).
GN NRP2 OR VEGF165R2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
RX MEDLINE-97470888; PubMed-9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III.";
RL Neuron 19:547-559(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A22).
RC TISSUE=Brain;
RX MEDLINE-98188099; PubMed-9529250;
RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an
RT isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745(1998).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE-20309748; PubMed-10748121;
RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
RT form of vascular endothelial growth factor (VEGF) and of placenta
RT growth factor-2, but only neuropilin-2 functions as a receptor for
RT the 145-amino acid form of VEGF.";
RL J. Biol. Chem. 275:18040-18045(2000).
CC -|- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC -|- SUBUNIT: NEUROPIILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH

```

NEUROPIILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN 3C RECEPTOR.

-|- SUBCELLULAR LOCATION: Type 1 membrane protein.

-|- ALTERNATIVE PRODUCTS: 3 isoforms; A22 (SHOWN HERE), A0 AND A17; ARE PRODUCED BY ALTERNATIVE SPLICING.

-|- SIMILARITY: BELONGS TO THE NEUROPIILIN FAMILY.

-|- SIMILARITY: CONTAINS 2 CUB DOMAINS.

-|- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.

-|- SIMILARITY: CONTAINS 1 MAM DOMAIN.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/submit/> or send an email to license@ebi.ac.uk).

EMBL; AF022859; AAC51788.1; -

EMBL; AF022860; AAC51789.1; -

EMBL; AF016098; AAC12922.1; -

HSSP; P12259; 1CZT.

Genew; HGNC:8005; NRP2.

MIM; 602070; -

InterPro; IPR000859; CUB_domain.

InterPro; IPR000421; FA58_C.

InterPro; IPR000998; MAM_domain.

Pfam; PF00431; CUB; 2.

Pfam; PF00629; MAM; 1.

Pfam; PF00754; F5_F8_type_C; 2.

SMART; SM00042; CUB; 2.

SMART; SM00231; FA58C; 2.

SMART; SM00137; MAM; 1.

PROSITE; PS01180; CUB; 2.

PROSITE; PS01285; FA58C_1; 2.

PROSITE; PS01286; FA58C_2; 2.

PROSITE; PS00060; MAM_2; 1.

Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;

Alternative splicing.

SIGNAL 1 20

CHAIN 21 931

DOMAIN 21 864

TRANSMEM 865 889

DOMAIN 890 931

DOMAIN 28 142

DOMAIN 149 267

DOMAIN 277 427

DOMAIN 434 592

DOMAIN 642 802

DOMAIN 671 674

DISULFID 28 55

DISULFID 83 105

DISULFID 149 175

DISULFID 208 230

DISULFID 277 427

DISULFID 434 592

CARBOHYD 152 152

CARBOHYD 157 157

CARBOHYD 629 629

CARBOHYD 839 839

VARSPLIC 809 813

VARSPLIC 809 830

CONFLICT 602 602

SEQUENCE 931 AA; 104830 MW; 270CBAE69A0A797C CRC64;

OR 22 (POTENTIAL).

NEUROPIILIN-2.

EXTRACELLULAR (POTENTIAL).

POTENTIAL.

CYTOPLASMIC (POTENTIAL).

CUB 1.

CUB 2.

F5/8 TYPE C 1.

F5/8 TYPE C 2.

MAM.

POLY-SER.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

BY SIMILARITY.

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

MISSING (IN ISOFORM A17).

MISSING (IN ISOFORM A0).

E -> K (IN REF. 1).

ALIGNMENT SCORES:

Pred. No.: 1.29e-20 Length: 931

Score: 398.50 Matches: 199

Percent Similarity: 36.37% Conservative: 112

Best Local Similarity: 23.27% Mismatches: 340

Query Match: 9.62% Indels: 204

DB: 1 Gaps: 36

US-10-060-830-1 (1-2280) x NRP2_HUMAN (1-931)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: January 21, 2003, 09:55:46 ; Search time 24.5702 Seconds
(without alignments)
5476.098 Million cell updates/sec

Title: US-10-060-830-1114

Perfect score: 653

Sequence: 1 MPLFLLLLVLLLEDDAGA.....TQEVSGAGRGCECDVFKREIL 653

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organella.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match %	Length	ID	Description
1	631	96.6	775	4 Q96PD2	Q96pd2 homo sapien
2	631	96.6	775	4 Q8TDX2	Q8tdx2 homo sapien
3	69	10.6	364	4 Q14089	Q14089 homo sapien
4	51	7.8	251	11 Q9D9K5	Q9d9k5 mus musculus
5	51	7.8	769	11 Q91ZV3	Q91zv3 mus musculus
6	44	6.7	769	11 Q91ZV2	Q91zv2 rattus norv
7	16	2.5	34	4 Q96IX0	Q96ix0 homo sapien
8	9	1.4	73	2 Q9EUA9	Q9eua9 helicobacte
9	9	1.4	73	2 Q9F6T0	Q9f6t0 helicobacte
10	9	1.4	73	2 Q9F6S9	Q9f6s9 helicobacte
11	9	1.4	73	2 Q9F6S8	Q9f6s8 helicobacte
12	9	1.4	155	4 Q9GPF81	Q9gpf81 homo sapien
13	9	1.4	237	8 Q8WEB2	Q8web2 mabuya dela
14	9	1.4	237	8 Q8WB8F	Q8wb8f mabuya dela
15	9	1.4	276	16 Q25256	Q25256 helicobacte
16	9	1.4	296	2 Q51756	Q51756 pseudomonas

17	9	1.4	332	8 Q9B179	Q9b179 mabuya dela
18	9	1.4	332	8 Q9B201	Q9b201 mabuya dela
19	9	1.4	332	8 Q9B200	Q9b200 mabuya dela
20	9	1.4	377	4 Q8TAG4	Q8tag4 homo sapien
21	9	1.4	728	16 Q9HYQ6	Q9hyq6 pseudomonas
22	9	1.4	1197	13 Q90478	Q90478 brachydanio
23	8	1.2	37	4 Q8TD81	Q8td81 homo sapien
24	8	1.2	71	10 Q8W483	Q8w483 arabidopsis
25	8	1.2	95	10 Q9SKJ8	Q9skj8 arabidopsis
26	8	1.2	95	10 Q90666	Q90666 arabidopsis
27	8	1.2	103	11 Q9CUJ3	Q9cu13 mus musculu
28	8	1.2	104	17 Q9UWS7	Q9uws7 sulfolobus
29	8	1.2	112	10 Q9FIG4	Q9fig4 arabidopsis
30	8	1.2	125	5 Q9UIP6	Q9uip6 caenorhabdi
31	8	1.2	139	16 Q92JV2	Q92jv2 rhizobium m
32	8	1.2	157	5 Q8T6R9	Q8t6r9 anopheles g
33	8	1.2	161	5 Q16655	Q16655 caenorhabdi
34	8	1.2	163	16 Q92PM3	Q92pm3 rhizobium m
35	8	1.2	201	17 Q58922	Q58922 pyrococcus
36	8	1.2	203	8 Q9MIB5	Q9mib5 mabuya eleg
37	8	1.2	222	16 Q8YN39	Q8yn39 anabaena sp
38	8	1.2	236	2 Q87318	Q87318 mycobacteri
39	8	1.2	245	11 Q9DIX2	Q9dix2 mus musculu
40	8	1.2	254	4 Q8WXV2	Q8wxv2 homo sapien
41	8	1.2	257	10 Q94CK4	Q94ck4 arabidopsis
42	8	1.2	258	16 Q9Z8D8	Q9z8d8 chlamydia p
43	8	1.2	271	10 Q8VZP2	Q8vzp2 arabidopsis
44	8	1.2	277	10 Q9FWJ5	Q9fmj5 arabidopsis
45	8	1.2	280	5 Q9Y166	Q9y166 drosophila

ALIGNMENTS

RESULT 1

Q96PD2 ID Q96PD2 PRELIMINARY; PRT; 775 AA.

AC Q96PD2; DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
DE ESDN.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury.";
RT J. Biol. Chem. 276:34105-34114(2001).
RL -!- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC EMBL; AF387547; AAL30178.1; .
DR InterPro; IPR000859; CUB.domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_typeC; 1.
DR PROSITE; PS01180; CUB; 1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;

Query Match 96.6%; Score 631; DB 4; Length 775;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 631; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 23 GKYCGLGLQMNHSIBSKGNEITLLFMSGIHVSGRGFLASYSVIDKQDLITCLDTASNELE 82
DB 145 GKYCGLGLQMNHSIBSKGNEITLLFMSGIHVSGRGFLASYSVIDKQDLITCLDTASNELE 204
QY 83 PEFKSKYCPAGCLLPFAEISGTIPHGYSRDSPLCMAGVHAGVVSNTLGGQISVWISKGIPY 142
DB 205 PEFKSKYCPAGCLLPFAEISGTIPHGYSRDSPLCMAGVHAGVVSNTLGGQISVWISKGIPY 264
QY 143 YESSLANNVTSVGHLSLSLETFKTSYGCYGLTGMESGVIADPQITASSVLEWTDHTGOEN 202
DB 265 YESSLANNVTSVGHLSLSLETFKTSYGCYGLTGMESGVIADPQITASSVLEWTDHTGOEN 324
QY 203 SWKPKKARLKKPGPPWAAFAFATDEYQWLQIDLNKKEKITGIITTTGSTMVHEHNYVSAYRIL 262
DB 325 SWKPKKARLKKPGPPWAAFAFATDEYQWLQIDLNKKEKITGIITTTGSTMVHEHNYVSAYRIL 384
QY 263 YSDGQKWTVYREPGEVDKIFQGNKDYHQDVRRNFPPIIARFIRVNPTQWQKIAMKM 322
DB 385 YSDGQKWTVYREPGEVDKIFQGNKDYHQDVRRNFPPIIARFIRVNPTQWQKIAMKM 444
QY 323 ELLGCOFTPKGRPKLTOPPPRNSNDLKNNTAPPKIAKGRAPKFTQLOPRSSNEFFPAQ 382
DB 445 ELLGCOFTPKGRPKLTOPPPRNSNDLKNNTAPPKIAKGRAPKFTQLOPRSSNEFFPAQ 504
QY 383 TEQTTASPDIRNTTVPNTVKDVALAALVLPVLMVLTLLILLYCAWHNRNKKKTEGT 442
DB 505 TEQTTASPDIRNTTVPNTVKDVALAALVLPVLMVLTLLILLYCAWHNRNKKKTEGT 564
QY 443 YDLPYWDRAKWKGMKQFLPAKAVDHEETPVRYSSSEVNHLSPREVTTVLQADSAYEYQ 502
DB 565 YDLPYWDRAKWKGMKQFLPAKAVDHEETPVRYSSSEVNHLSPREVTTVLQADSAYEYQ 624
QY 503 LVGGIVGLTHORSTFKPEGKEAGYADLDPNYSPQGEVYHAYAEPLITGPEYATPIIMD 562
DB 625 LVGGIVGLTHORSTFKPEGKEAGYADLDPNYSPQGEVYHAYAEPLITGPEYATPIIMD 684
QY 563 MSGHPTTSVGPQSTSTFRKATGNQPPPLVGTNTLLSRDSCSSAAQYDTPKAGKPLPA 622
DB 685 MSGHPTTSVGPQSTSTFRKATGNQPPPLVGTNTLLSRDSCSSAAQYDTPKAGKPLPA 744
QY 623 PDELVIQVQSTQEVSGAGRGECDFVFEIL 653
DB 745 PDELVIQVQSTQEVSGAGRGECDFVFEIL 775

```

RESULT 2

```

Q8TDX2
ID Q8TDX2 PRELIMINARY; PRT; 775 AA.
AC Q8TDX2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CLCPI.
GN CLCPI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Koshikawa K., Osada H., Kozaki K., Konishi H., Masuda A.,
RA Tatamatsu Y., Mitsuomi T., Nakao A., Takahashi T.;
RT "Significant up-regulation of a novel gene, CLCPI, in a highly
RT metastatic lung cancer subline as well as in lung cancers in vivo.";
RL Oncogene 0:0-0(2002).
DR EMBL; AB073146; BAB91138.1;
SQ SEQUENCE 775 AA; 85073 MW; 6007223B21BE5A42 CRC64;

```

Query Match 96.6%; Score 631; DB 4; Length 775;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 631; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 3

```

Q14089
ID Q14089 PRELIMINARY; PRT; 364 AA.
AC Q14089;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.0 kDa protein (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Shibata T.;
RT "unpublished.";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; D29810; BAA18909.1;
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FAS8_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
KW Hypothetical protein.
FT NON_TER 1

```

SQ SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;

Query Match 10.6%; Score 69; DB 4; Length 364;
Best Local Similarity 100.0%; Pred. No. 1.6e-64;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 90 PAGCLLPAEISGTPHGYRDSPLCMAGVHAGVYVNTLGGQISVWISKGPYYESSLAN 149
|||||
Db 144 PAGCLLPAEISGTPHGYRDSPLCMAGVHAGVYVNTLGGQISVWISKGPYYESSLAN 203
|||||
Oy 150 NVTWVGHL 158
|||||
Db 204 NVTWVGHL 212
|||||

RESULT 4

Q9D9K5 ID Q9D9K5 PRELIMINARY; PRT; 251 AA.
AC Q9D9K5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 1700055P2IRik protein.
GN ESDN OR 1700055P2IRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=TESTIS;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Willming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK006805; BAB24750.1; -.
DR MGD; MGI:1920629; ESDN.
SQ SEQUENCE 251 AA; 27428 MW; F86E0AA15EF51AE9 CRC64;

Query Match 7.8%; Score 51; DB 11; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.8e-45;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 486 REVTTVLQADSAYAQPLVGGVIGTTHORSTFKPEEGKEAGYADLPYNP 536
|||||
Db 85 REVTTVLQADSAYAQPLVGGVIGTTHORSTFKPEEGKEAGYADLPYNP 135
|||||

RESULT 5

Q91ZV3 ID Q91ZV3 PRELIMINARY; PRT; 769 AA.

AC Q91ZV3;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Endothelial and smooth muscle cell-derived neuropilin-like

DE protein.
GN ESDN.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=ICR;

RC PubMed=11447234;

RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,

RA Matsumori A., Sasayama S., Honjo T., Tashiro K.,

RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular

RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is

RT Up-regulated after Vascular Injury";

RL J. Biol. Chem. 276.34105-34114(2001).

CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL

CC REGULATORS.

DR EMBL; AF387548; AAL30179.1; -.
DR MGD; MGI:1920629; ESDN.

DR InterPro; IPR000859; CUB_domain.

DR InterPro; IPR000421; FA58_C.

DR InterPro; IPR004043; LCCL_dom.

DR Pfam; PF00431; CUB; 1.

DR Pfam; PF00754; F5_F8_type_C; 1.

DR PROSITE; PS01180; CUB; 1.

KW DNA-binding; Transcription regulation.

SQ SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;

Query Match 7.8%; Score 51; DB 11; Length 769;

Best Local Similarity 100.0%; Pred. No. 5.1e-45;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 486 REVTTVLQADSAYAQPLVGGVIGTTHORSTFKPEEGKEAGYADLPYNP 536
|||||

Db 603 REVTTVLQADSAYAQPLVGGVIGTTHORSTFKPEEGKEAGYADLPYNP 653
|||||

RESULT 6

Q91ZV2 ID Q91ZV2 PRELIMINARY; PRT; 769 AA.

AC Q91ZV2;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Endothelial and smooth muscle cell-derived neuropilin-like

DE protein.

GN ESDN.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY;

RX PubMed=11447234;

RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,

RA Matsumori A., Sasayama S., Honjo T., Tashiro K.,

RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular

RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is

RT Up-regulated after Vascular Injury";

RL J. Biol. Chem. 276.34105-34114(2001).

CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL

CC REGULATORS.

DR EMBL; AF387549; AAL30180.1; -.
DR InterPro; IPR000859; CUB_domain.

DR InterPro; IPR000421; FA58_C.

DR InterPro; IPR004043; LCCL_dom.

DR Pfam; PF00431; CUB; 1.

DR Pfam; PF00754; F5_F8_type_C; 1.

DR PROSITE; PS01180; CUB; 1.

KW DNA-binding; Transcription regulation.

SQ SEQUENCE 769 AA; 83915 MW; 5B7936C8FA063A93 CRC64;

Query Match 6.7%; Score 44; DB 11; Length 769;
Best Local Similarity 100.0%; Pred. No. 1.5e-37; Indels 0; Gaps 0;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 482 HLSPREVTTVLQADSAYEYQPLVGGIVGTLHQRSTFKPEGKEA 525
|||||
DB 599 HLSPREVTTVLQADSAYEYQPLVGGIVGTLHQRSTFKPEGKEA 642
|||||

RESULT 7

ID Q96IX0 PRELIMINARY; PRT; 34 AA.
AC Q96IX0;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Unknown (Protein for IMAGE:433376) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007117; AAH07117.1; -;
FT NON_TER 1
SQ SEQUENCE 34 AA; 3533 MW; E6EA7A2239151709 CRC64;

Query Match 2.5%; Score 16; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 7.1e-09; Indels 0; Gaps 0;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LLLVLLLEDDAGAAQ 22
|||||
DB 1 LLLVLLLEDDAGAAQ 16
|||||

RESULT 8

ID Q9EUA9 PRELIMINARY; PRT; 73 AA.
AC Q9EUA9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HP0519-like protein (Fragment).
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR41, CR45, AND CR10;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and gene expression of the cag pathogenicity island in
RT Helicobacter pylori strains isolated from gastric carcinoma and
RT gastritis patients in Costa Rica";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF289398; AAG09847.1; -;
DR EMBL; AF289398; AAG09847.1; -;
DR EMBL; AF289393; AAG09837.1; -;
DR EMBL; AF289397; AAG09845.1; -;
FT NON_TER 73
SQ SEQUENCE 73 AA; 8853 MW; 1538A4F55832FC54 CRC64;

Query Match 1.4%; Score 9; DB 2; Length 73;
Best Local Similarity 100.0%; Pred. No. 0.42; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLVLLVLL 13
|||||

Db 16 LLLVLLVLL 24
|||||

RESULT 9

ID Q9F6T0 PRELIMINARY; PRT; 73 AA.
AC Q9F6T0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HP0519-like protein (Fragment).
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR46;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL; AF289394; AAG09839.1; -;
FT NON_TER 73
SQ SEQUENCE 73 AA; 8819 MW; 1BD644F55838F654 CRC64;

Query Match 1.4%; Score 9; DB 2; Length 73;
Best Local Similarity 100.0%; Pred. No. 0.42; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLVLLVLL 13
|||||
DB 16 LLLVLLVLL 24
|||||

RESULT 10

ID Q9F6S9 PRELIMINARY; PRT; 73 AA.
AC Q9F6S9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HP0519-like protein (Fragment).
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR47;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL; AF289395; AAG09841.1; -;
FT NON_TER 73
SQ SEQUENCE 73 AA; 8818 MW; 07B4D3B81224EA5E CRC64;

Query Match 1.4%; Score 9; DB 2; Length 73;
Best Local Similarity 100.0%; Pred. No. 0.42; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLVLLVLL 13
|||||
DB 16 LLLVLLVLL 24
|||||

```

RESULT 11
Q9F6S8
ID Q9F6S8 PRELIMINARY; PRT; 73 AA.
AC Q9F6S8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HP0519-like protein (Fragment).
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR39;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica";
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL; AF289396; AAG09843.1; -
FT NON_TER 73
SQ SEQUENCE 73 AA; 8817 MW; 07B4D918B824EA5E CRC64;

Query Match 1.4%; Score 9; DB 2; Length 73;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLLLL 13
|
Db 16 LLLLLLLLL 24

RESULT 12
Q96P81
ID Q96P81 PRELIMINARY; PRT; 155 AA.
AC Q96P81;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PTPRE (Fragment).
GN PTPRE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Zhao Y., Sun D., Dai S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF406557; AAL01375.1; -
FT NON_TER 155
SQ SEQUENCE 155 AA; 17359 MW; 8A7DB29677F06E00 CRC64; --

Query Match 1.4%; Score 9; DB 4; Length 155;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLLLL 13
|
Db 56 LLLLLLLLL 64

RESULT 13
Q8WEB2
ID Q8WEB2 PRELIMINARY; PRT; 237 AA.
AC Q8WEB2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

```

```

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cytochrome b (Fragment).
GN CYTB.
OS Mabuva delalandii.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Scincomorpha; Scincoidae;
OC Scincidae; Mabuva.
OX NCBI_TaxID=149020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=67MDEST;
RA Carranza S., Arnold N., Mateo J., Lopez-Jurado L.;
RT "Parallel gigantism and complex colonization patterns in Cape Verde
RT scincid lizards Mabuva and Macroscincus (Reptilia: Scincidae) revealed
RT by mitochondrial DNA sequences.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280320; AAL55225.1; -
DR InterPro; IPR000179; Cyt_b_b6.
DR Pfam; PF00033; cytochrome_b_N; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
KW Mitochondrion.
FT NON_TER 1
FT NON_SEQUENCE 237
SQ SEQUENCE 237 AA; 26439 MW; 73A5BADDDBA7F7D9A CRC64;

Query Match 1.4%; Score 9; DB 8; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LLLLLLLLL 14
|
Db 202 LLLLLLLLL 210

RESULT 14
Q8W8F8
ID Q8W8F8 PRELIMINARY; PRT; 237 AA.
AC Q8W8F8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cytochrome b (Fragment).
GN CYTB.
OS Mabuva delalandii.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Scincomorpha; Scincoidae;
OC Scincidae; Mabuva.
OX NCBI_TaxID=149020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RA Carranza S., Arnold N., Mateo J., Lopez-Jurado L.;
RT "Parallel gigantism and complex colonization patterns in Cape Verde
RT scincid lizards Mabuva and Macroscincus (Reptilia: Scincidae) revealed
RT by mitochondrial DNA sequences.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280315; AAL55220.1; -
DR EMBL; AF280316; AAL55221.1; -
DR EMBL; AF280317; AAL55221.1; -
DR EMBL; AF280319; AAL55224.1; -
DR EMBL; AF280321; AAL55226.1; -
DR EMBL; AF280322; AAL55227.1; -
DR EMBL; AF280323; AAL55228.1; -
DR EMBL; AF280324; AAL55229.1; -
DR EMBL; AF280325; AAL55230.1; -
DR EMBL; AF280326; AAL55231.1; -
DR EMBL; AF280327; AAL55232.1; -
DR InterPro; IPR000179; Cyt_b_b6.
DR Pfam; PF00033; cytochrome_b_N; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
KW Mitochondrion.

```

```
FT NON_TER 1
FT NON_TER 237
SQ SEQUENCE 237 AA; 26455 MW; 7297BADD8A7F7D84 CRC64;

Query Match 1.4%; Score 9; DB 8; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LLLLLLLLLL 14
   |
Db 202 LLLLLLLLLL 210

RESULT 15
O25256 PRELIMINARY; PRT; 276 AA.
AC O25256;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein HP0519.
GN HP0519.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RA "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
DR EMBL; AE000566; AAD07587.1; -.
DR TIGR; HP0519; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 276 AA; 31527 MW; 7B987D1860EF9743 CRC64;

Query Match 1.4%; Score 9; DB 16; Length 276;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLLLLL 13
   |
Db 16 LLLLLLLLLL 24
```

Search completed: January 21, 2003, 09:59:31
Job time : 26.5702 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:55:15 ; Search time 10.8676 Seconds
(without alignments)
1767.934 Million cell updates/sec

Title: US-10-060-830-1114
Perfect score: 653
Sequence: 1 MFLFLLLVLLVLLLEDDAGA.....TOEVSGAGRGCDVFEKIL 653

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	1.2	113	4	US-08-816-772-9
2	8	1.2	114	1	US-08-352-324A-3
3	8	1.2	114	2	US-08-862-607-3
4	8	1.2	114	2	US-08-468-819-4
5	8	1.2	114	3	US-08-203-235-3
6	8	1.2	114	4	US-08-679-493A-149
7	8	1.2	114	5	PCT-US95-16144-3
8	8	1.2	117	4	US-09-134-001C-5606
9	8	1.2	349	4	US-09-032-523-3
10	8	1.2	358	2	US-08-748-485-6
11	8	1.2	396	1	US-08-208-007A-13
12	8	1.2	396	4	US-09-032-523-9
13	8	1.2	396	4	US-08-915-095A-13
14	8	1.2	396	4	US-08-798-096-13
15	8	1.2	396	4	US-08-798-095A-13
16	8	1.2	480	2	US-08-828-488-8
17	8	1.2	480	4	US-09-299-689A-8
18	8	1.2	574	2	US-08-756-317-4
19	8	1.2	600	2	US-08-756-317-3
20	8	1.2	638	2	US-08-756-317-2
21	8	1.2	1503	4	US-08-976-255-14
22	7	1.1	15	4	US-09-336-536-5
23	7	1.1	19	3	US-08-448-194-12
24	7	1.1	19	4	US-08-867-921-12
25	7	1.1	30	4	US-08-448-489-7
26	7	1.1	100	3	US-08-476-376-2
27	7	1.1	108	4	US-08-654-482-4

28	7	1.1	109	4	US-09-071-035-194	Sequence 194, Appl
29	7	1.1	135	4	US-09-388-143-69	Sequence 69, Appl
30	7	1.1	144	2	US-08-341-843B-27	Sequence 27, Appl
31	7	1.1	144	2	US-08-427-497B-32	Sequence 32, Appl
32	7	1.1	150	2	US-08-867-676-3	Sequence 3, Appl
33	7	1.1	151	2	US-08-620-694A-8	Sequence 8, Appl
34	7	1.1	151	3	US-09-034-810-6	Sequence 6, Appl
35	7	1.1	151	3	US-09-022-255-8	Sequence 8, Appl
36	7	1.1	151	3	US-09-022-255-8	Sequence 8, Appl
37	7	1.1	151	3	US-08-685-239-6	Sequence 8, Appl
38	7	1.1	151	3	US-09-022-253-8	Sequence 8, Appl
39	7	1.1	151	3	US-09-022-260-8	Sequence 8, Appl
40	7	1.1	151	4	US-09-022-259-8	Sequence 8, Appl
41	7	1.1	151	4	US-09-022-257-8	Sequence 8, Appl
42	7	1.1	151	4	US-08-432-994A-4	Sequence 4, Appl
43	7	1.1	161	2	US-08-386-819A-25	Sequence 25, Appl
44	7	1.1	161	3	US-08-960-357-25	Sequence 25, Appl
45	7	1.1	163	2	US-08-867-676-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-816-772-9
; Sequence 9, Application US/08816772
; Patent No. 6410268
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
; APPLICANT: LI, HAODONG
; APPLICANT: SU, JEFFREY
; TITLE OF INVENTION: CHEMOKINE ALPHA-3
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,772
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0980001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-816-772-9

Query Match 1.2%; Score 8; DB 4; Length 113;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LLVLLLLL 15
|||||||
Db 21 LLVLLLLL 28

SEQUENCE CHARACTERISTICS:
LENGTH: 114 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-468-819-4

Query Match 1.2%; Score 8; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 LLVLLLLL 15
| | | | |
Db 21 LLVLLLLL 28

RESULT 5

US-09-203-235-3
Sequence 3, Application US/09203235
Patent No. 6071701

GENERAL INFORMATION:
APPLICANT: Guegler, Karl J.
APPLICANT: Hawkins, Phillip R.
APPLICANT: Wilde, Craig G.
APPLICANT: Seilhamer, Jeffrey J.
TITLE OF INVENTION: A NOVEL CHEMOKINE EXPRESSED IN
TITLE OF INVENTION: INFLAMED ADENOID, ITS PRODUCTION AND USES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/203,235
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/862,607
FILING DATE: 23-MAY-1997
APPLICATION NUMBER: 08/352,324
FILING DATE: 07-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33,954
REFERENCE/DOCKET NUMBER: PF-0025 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 114 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-203-235-3

Query Match 1.2%; Score 8; DB 3; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 LLVLLLLL 15
| | | | |
Db 21 LLVLLLLL 28

RESULT 6

US-08-679-493A-149
Sequence 149, Application US/08679493A
Patent No. 6303295

GENERAL INFORMATION:
APPLICANT: Taylor, Ethan W.
TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
FILE REFERENCE: 55-95
CURRENT APPLICATION NUMBER: US/08/679,493A
CURRENT FILING DATE: 1996-07-12
PRIOR APPLICATION NUMBER: 60/001203
PRIOR FILING DATE: 1995-07-14
PRIOR APPLICATION NUMBER: 60/003,112
PRIOR FILING DATE: 1995-09-01
NUMBER OF SEQ ID NOS: 216
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 149
LENGTH: 114
TYPE: PRT
ORGANISM: Homo sapiens
US-08-679-493A-149

Query Match 1.2%; Score 8; DB 4; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 LLVLLLLL 15
| | | | |
Db 21 LLVLLLLL 28

RESULT 7

PCT-US95-16144-3
Sequence 3, Application PC/TUS9516144
GENERAL INFORMATION:

APPLICANT: INCYTE PHARMACEUTICALS, INC.
TITLE OF INVENTION: A NOVEL CHEMOKINE EXPRESSED IN INFLAMED
TITLE OF INVENTION: ADENOID, ITS PRODUCTION AND USES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16144
FILING DATE: 07-DEC-1995
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,324
FILING DATE: 07-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: LUTHER, BARBARA J.
REGISTRATION NUMBER: 33954
REFERENCE/DOCKET NUMBER: PF-0025 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 114 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

us-10-060-830-1114.oli.ra1

Wed Jan 22 14:57:25 2003

```

; MOLECULE TYPE: protein
; PCT-US95-16144-3
;
; Query Match 1.2% Score 8; DB 5; Length 114;
; Best Local Similarity 100.0%; Pred. No. 7.1;
; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 8 LVLVLLL 15
; Db 21 LVLVLLL 28
;
; RESULT 8
; US-09-134-001C-5606
; Sequence 5606, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-08-14
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5606
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5606
;
; Query Match 1.2% Score 8; DB 4; Length 117;
; Best Local Similarity 100.0%; Pred. No. 7.3;
; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 3 LFLVLLL 10
; Db 32 LFLVLLL 39
;
; RESULT 9
; US-09-032-523-3
; Sequence 3, Application US/09032523
; Patent No. 623454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; APPLICANT: Baugh, Mariah
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,523
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
;
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0159 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0479 US
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PANTUT01
; CLONE: 1515165
; US-09-032-523-3
;
; Query Match 1.2% Score 8; DB 4; Length 349;
; Best Local Similarity 100.0%; Pred. No. 19;
; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 5 LLLLLLV 12
; Db 5 LLLLLLV 12
;
; RESULT 10
; US-08-748-485-6
; Sequence 6, Application US/08748485
; Patent No. 5817480
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Guegler, Karl J.
; APPLICANT: Goli, Surya K.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: NOVEL HISTAMINE H2 RECEPTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,485
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0159 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;

```

IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 236184
US-08-748-485-6

Query Match 1.2%; Score 8; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 418 VLTTLILI 425
Db 24 VLTTLILI 31

RESULT 11

US-08-208-007A-13
Sequence 13, Application US/08208007A
Patent No. 5501969
GENERAL INFORMATION:
APPLICANT: HASTINGS, ET AL.
TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/208,007A
FILING DATE: March 8, 1994.
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: NO. 5501969e
FILING DATE: NO. 5501969e
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-208-007A-13

Query Match 1.2%; Score 8; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLLVL 12
Db 5 LLLLLLVL 12

RESULT 12

US-09-032-523-9
Sequence 9, Application US/09032523
Patent No. 6232454
GENERAL INFORMATION:
APPLICANT: Bandman, Olga

APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl
APPLICANT: Baugh, Mariah
TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0479 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 181994
US-09-032-523-9

Query Match 1.2%; Score 8; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLLVL 12
Db 5 LLLLLLVL 12

RESULT 13
US-08-915-095A-13
Sequence 13, Application US/08915095A
Patent No. 6383793
GENERAL INFORMATION:
APPLICANT: Hastings, et al.
TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: PF107D4
CURRENT APPLICATION NUMBER: US/08/915,095A
CURRENT FILING DATE: 1997-08-20
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 396
TYPE: PRT
ORGANISM: Homo sapiens
US-08-915-095A-13

Query Match 1.2%; Score 8; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 LLLLLVLL 12
 Db 5 LLLLLVLL 12

RESULT 14
 US-08-798-096-13
 ; Sequence 13, Application US/08798096
 ; Patent No. 6387682
 ; GENERAL INFORMATION:
 ; APPLICANT: Hastings, et al.
 ; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
 ; FILE REFERENCE: PF107D2
 ; CURRENT APPLICATION NUMBER: US/08/798,096
 ; CURRENT FILING DATE: 1997-02-12
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 13
 ; LENGTH: 396
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-08-798-096-13

Query Match 1.2%; Score 8; DB 4; Length 396;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 LLLLLVLL 12
 Db 5 LLLLLVLL 12

RESULT 15
 US-08-798-095A-13
 ; Sequence 13, Application US/08798095A
 ; Patent No. 6423507
 ; GENERAL INFORMATION:
 ; APPLICANT: Hastings, et al.
 ; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
 ; FILE REFERENCE: PF107D3
 ; CURRENT APPLICATION NUMBER: US/08/798,095A
 ; CURRENT FILING DATE: 1997-02-12
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 13
 ; LENGTH: 396
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-08-798-095A-13

Query Match 1.2%; Score 8; DB 4; Length 396;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 LLLLLVLL 12
 Db 5 LLLLLVLL 12

Search completed: January 21, 2003, 09:58:22
 Job time : 11.8676 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:50:02 ; Search time 7.32283 Seconds
(without alignments)
1628.756 Million cell updates/sec

Title: US-10-060-830-4

Perfect score: 99

Sequence: 1 aaacagaagaaaaaactg.....tatgcagacctatgacctta 300

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 120991 seqs, 19878514 residues

Word size: 1

Total number of hits satisfying chosen parameters: 218156

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US10060830/runat_16012003.092704_1550/app_query.fasta_1.1109
-DB=publishedApplications_AA -QFMT=fastan -SUFFIX=olin2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10060830 -ECGN_1_1_4 -runat_16012003.092704_1550
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

PublishedApplications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	8	8.2	242	10	US-09-864-761-36180
c 2	8	8.2	447	9	US-09-870-759-50
c 3	7	7.2	30	10	US-09-987-190-15
c 4	7	7.1	31	10	US-09-864-761-41712
					Sequence 36180, A
					Sequence 50, Appl
					Sequence 15, Appl
					Sequence 41712, A

ALIGNMENTS

RESULT 1

US-09-864-761-36180 Application US/09864761
; Sequence 36180, Appl

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F

; FILE REFERENCE: Acomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864.761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

Sequence 397, App
Sequence 1863, Ap
Sequence 104, App
Sequence 1195, Ap
Sequence 2, Appl
Sequence 190, App
Sequence 5, Appl
Sequence 5, Appl
Sequence 10, Appl
Sequence 12, Appl
Sequence 18, Appl
Sequence 18, Appl
Sequence 18, Appl
Sequence 46, Appl
Sequence 36, Appl
Sequence 53, Appl
Sequence 53, Appl
Sequence 2, Appl
Sequence 37, Appl
Sequence 4, Appl
Sequence 12, Appl
Sequence 48, Appl
Sequence 49, Appl
Sequence 51, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 61, Appl
Sequence 2, Appl
Sequence 12, Appl
Sequence 2, Appl
Sequence 218, App
Sequence 6, Appl
Sequence 803, App
Sequence 803, App
Sequence 803, App
Sequence 803, App
Sequence 803, App
Sequence 803, App
Sequence 11, Appl
Sequence 802, App
Sequence 802, App
Sequence 802, App

```
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36180
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022333.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
;
US-09-864-761-36180
Alignment Scores:
Pred. No.: 2.59 Length: 242
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 10 Gaps: 0
US-10-060-830-4 (1-300) x US-09-864-761-36180 (1-242)
QY 281 AGCTGCTTCTTTCTCTTCTG 258
Db 52 SerLeuLeuPheLeuLeu 59
|||||
RESULT 2
US-09-870-759-50
; Sequence 50, Application US/09870759
; Patent No. US20020177551A1
; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 870759
; CURRENT APPLICATION NUMBER: US/09/870,759
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: US 60/208,128
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 447
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-09-870-759-50
Alignment Scores:
Pred. No.: 2.49 Length: 447
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 9 Gaps: 0
US-10-060-830-4 (1-300) x US-09-870-759-50 (1-447)
QY 281 AGCTGCTTCTTTCTCTTCTG 258
Db 273 SerLeuLeuPheLeuLeu 280
|||||
RESULT 3
US-09-987-190-15
; Sequence 15, Application US/09987190
; Patent No. US20020058293A1
; GENERAL INFORMATION:
; APPLICANT: TAKESAKO, Kazutoh
; APPLICANT: MIZUTANI, Shigetoshi
; APPLICANT: ENDO, Masahiro
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: FUNGAL ANTIGENS AND PROCESS FOR PRODUCING THE SAME
; FILE REFERENCE: 1422-0502p
; CURRENT APPLICATION NUMBER: US/09/987,190
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/262,856
; PRIOR FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-987-190-15
Alignment Scores:
Pred. No.: 31.5 Length: 30
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 10 Gaps: 0
US-10-060-830-4 (1-300) x US-09-987-190-15 (1-30)
QY 107 CATGGTCCACTGCTTTTCGAG 87
Db 24 HisGlyProLeuLeuLeuGln 30
|||||
RESULT 4
US-09-864-761-41712
; Sequence 41712, Application US/09864761
; Patent No. US2002048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
```

```

; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 397
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-397

Alignment Scores:
Pred. No.: 30.3 Length: 57
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-4 (1-300) x US-09-764-870-397 (1-57)

QY 281 AGCCTGCTCTCTTTTCCTTCTT 261
|||||
Db 49 SerLeuLeuPheLeuLeu 55

RESULT 6
US-09-764-877-1863
; Sequence 1863, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1863
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (47)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (68)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (95)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1863

Alignment Scores:
Pred. No.: 29.1 Length: 105
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-4 (1-300) x US-09-764-877-1863 (1-105)

QY 290 GGCTGTCATGCGCTTCTT 270
|||||
Db 56 GlyLeuHisSerLeuLeu 62

RESULT 7
US-10-028-072-104
; Sequence 104, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

```

APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang
TITLE OF INVENTION:
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/028,072
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059836
PRIOR FILING DATE: 1997-09-24
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062285
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062814
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/062816
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063045
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063082
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/063127
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063327
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063329
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063550
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063561
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063704
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063733
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063735
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063738
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063755
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064248
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/064809
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065846
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066453
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066511
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069212
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069278
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069334
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069694
PRIOR FILING DATE: 1997-12-16
PRIOR APPLICATION NUMBER: 60/072320
PRIOR FILING DATE: 1998-01-23
PRIOR APPLICATION NUMBER: 60/073612
PRIOR FILING DATE: 1998-02-04
PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-02-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081695
PRIOR FILING DATE: 1998-04-14
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085149
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13

; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086414
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086430
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088730
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088741
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090538
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07

Alignment Scores:
Pred. No.: 27.8 Length: 212
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 9 Gaps: 0

US-10-060-830-4 (1-300) x US-10-028-072-104 (1-212)

Oy 278 CTGCTCTTTTCCTCTCTG 258
|||||
Db 13 LeuLeuLeuPheLeuLeu 19

RESULT 8
US-09-925-301-1195
; Sequence 1195, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106

; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1195
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (245)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (246)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (257)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (266)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1195

Alignment Scores:
Pred. No.: 27.4 Length: 269
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-4 (1-300) x US-09-925-301-1195 (1-269)

Oy 227 GGTACACTTCATCAAGATCT 247
|||||
Db 145 GlyThrLeuHisGlnArgSer 151

RESULT 9
US-09-970-989-2
; Sequence 2, Application US/09970989
; Patent No. US20020156262A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, David W.
; APPLICANT: ADOUREL, Daniel
; APPLICANT: HOLLENBACK, David
; TITLE OF INVENTION: MAMMALIAN LYOPHOSPHATIDIC ACID ACYLTRANSFERASE
; FILE REFERENCE: 077319/0151
; CURRENT APPLICATION NUMBER: US/09/970,989
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/215,252
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: US 08/618,651
; PRIOR FILING DATE: 1996-03-19
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-970-989-2

Alignment Scores:
Pred. No.: 27.3 Length: 283
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 9 Gaps: 0

US-10-060-830-4 (1-300) x US-09-970-989-2 (1-283)

QY 278 CTGCTCTTTCTCTCTCTG 258
Db 12 LeuLeuLeuPheLeuLeu 18

RESULT 10
US-09-800-729-190
; Sequence 190, Application US/09800729
; Patent No. US20020068319A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCI/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 190
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-729-190

Alignment Scores:
Pred. No.: 27.2 Length: 310
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-4 (1-300) x US-09-800-729-190 (1-310)

QY 276 GCTCTTTCTCTCTCTG 256
Db 59 AlaserPheProserGly 65

RESULT 11
US-09-735-169A-5
; Sequence 5, Application US/09735169A
; Patent No. US20020028483A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; TITLE OF INVENTION: NOVEL POTASSIUM CHANNEL MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-074DV1
; CURRENT APPLICATION NUMBER: US/09/735,169A
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 09/259,951
; PRIOR FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-735-169A-5

Alignment Scores:
Pred. No.: 27 Length: 332
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-4 (1-300) x US-09-735-169A-5 (1-332)

QY 278 CTGCTTTCTCTCTCTG 258

Db 189 LeuLeuLeuPheLeuLeu 195

RESULT 12
US-09-735-171A-5
; Sequence 5, Application US/09735171A
; Patent No. US20020045209A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; TITLE OF INVENTION: NOVEL POTASSIUM CHANNEL MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-074DV2
; CURRENT APPLICATION NUMBER: US/09/735,171A
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 09/259,951
; PRIOR FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-735-171A-5

Alignment Scores:
Pred. No.: 27 Length: 332
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-4 (1-300) x US-09-735-171A-5 (1-332)

QY 278 CTGCTTTCTCTCTCTG 258

Db 189 LeuLeuLeuPheLeuLeu 195

RESULT 13
US-09-852-156-10
; Sequence 10, Application US/09852156
; Patent No. US20020076694A1
; GENERAL INFORMATION:
; APPLICANT: Littman, Dan R.
; Deng, Hongkul
; Unutmaz, Derya
; Ramani, Vineet N.K.
; TITLE OF INVENTION: NOVEL ALTERNATIVE G-COUPLED RECEPTORS
; ASSOCIATED WITH RETROVIRAL ENTRY INTO CELLS, METHODS OF
; IDENTIFYING THE SAME, AND DIAGNOSTIC AND THERAPEUTIC USE
; THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/852,156
; FILING DATE: 09-May-2001
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1049-1-009NCIP
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 360 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: African Green Monkey
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-852-156-10
Alignment Scores:
Pred. No.: 26.9 Length: 360
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 10 Gaps: 0
US-10-060-830-4 (1-300) x US-09-852-156-10 (1-360)
Qy 256 ACCAGAAGAGGAAGAAGC 276
Db 351 ThrArgArgLysArgSer 357
RESULT 14
US-09-852-156-12
Sequence 12, Application US/09852156
Patent No. US200200766941
GENERAL INFORMATION:
APPLICANT: Littman, Dan R.
Deng, Hongkui
Unutmaz, Derya
Ramani, Vineet N.K.
TITLE OF INVENTION: NOVEL ALTERNATIVE G-COUPLED RECEPTORS
ASSOCIATED WITH RETROVIRAL ENTRY INTO CELLS, METHODS OF
IDENTIFYING THE SAME, AND DIAGNOSTIC AND THERAPEUTIC USES
THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/852,156
FILING DATE: 09-May-2001
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1049-1-009NCIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 360 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: pigtail macaque
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-852-156-12
Alignment Scores:
Pred. No.: 26.9 Length: 360
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 10 Gaps: 0
US-10-060-830-4 (1-300) x US-09-852-156-12 (1-360)
Qy 256 ACCAGAAGAGGAAGAAGC 276
Db 351 ThrArgArgLysArgSer 357
RESULT 15
US-09-813-398-18
Sequence 18, Application US/09813398
Patent No. US20020169292A1
GENERAL INFORMATION:
APPLICANT: Bruce D. Weintraub
APPLICANT: Mariusz W. Szkudlinski
APPLICANT: University of Maryland
TITLE OF INVENTION: CYSTINE KNOT GROWTH FACTOR MUTANTS
FILE REFERENCE: UOFMD.003C1
CURRENT APPLICATION NUMBER: US/09/813,398
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: PCT/US99/05908
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: PCT/US98/19772
PRIOR FILING DATE: 1998-09-22
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 367
TYPE: PRT
ORGANISM: HOMO SAPIEN
US-09-813-398-18
Alignment Scores:
Pred. No.: 26.9 Length: 367
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 9 Gaps: 0
US-10-060-830-4 (1-300) x US-09-813-398-18 (1-367)
Qy 278 CTGCTTCCTTTCCCTTCCTTCG 258
Db 6 LeuLeuLeuPheLeuLeuLeu 12
Search completed: January 21, 2003, 14:55:25
Job time : 8.32283 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 02:16:00 ; Search time 57.2481 seconds
(without alignments)
10613.845 Million cell updates/sec

Title: US-10-060-830-1
Perfect score: 4143
Sequence: 1 gccgcgccccgcctgggc.....tttgaagatgatgtgcttt 2280

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0.

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2.1/USPNO_spool/US10060830/runat_16012003_091238_26833/app_query.fasta_1.6933
-DB-A_Geneseq_101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830 -CGCN_1.1_230 @runat_16012003_091238_26833 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3572	86.2	669	21	AA170539 Human Factor 8 Hom
2	2931	70.7	583	22	AA175450 Human colon cancer
3	1841.5	44.4	385	23	AA122716 Human neuropilin-H
4	1841.5	44.4	385	23	AA179460 Human Neuropilin-H
5	1841.5	44.4	385	23	AA179386 Novel human protei
6	1750.5	42.3	365	23	AA122721 Human neuropilin-H
7	1042	25.2	715	22	AA100670 Human TANGO 229 po
8	1039	25.1	197	22	AA101298 Novel human diagno
9	926	22.4	586	22	AA100629 Novel human protei
10	924.5	22.3	539	22	AA100630 Novel human protei
11	860.5	20.8	487	22	AA100628 Novel human protei
12	574.5	13.9	503	21	AA119126 Polypeptide isolat
13	540.5	13.0	398	23	AA122715 Human neuropilin-H
14	540.5	13.0	398	23	AA179459 Human Neuropilin-H
15	421.5	10.2	925	20	AA196308 Neuropilin-2. Rat
16	417	10.1	957	22	AA125770 Human protein sequ
17	414.5	10.0	901	20	AA196256 Mouse semaphorin r
18	414.5	10.0	906	20	AA196257 Mouse semaphorin r
19	414.5	10.0	909	20	AA196249 Rat semaphorin rec
20	412.5	10.0	909	20	AA196250 Mouse semaphorin r
21	412.5	10.0	909	21	AA124214 Mouse soluble neur
22	412.5	10.0	914	20	AA196251 Mouse semaphorin r
23	412.5	10.0	926	20	AA196252 Mouse semaphorin r
24	412.5	10.0	931	20	AA196253 Mouse semaphorin r
25	412	9.9	75	22	AA124217 Peptide #9723 enco
26	412	9.9	75	22	AA163102 Human brain expres
27	412	9.9	75	22	AA175913 Human bone marrow
28	412	9.9	75	22	AA136024 Peptide #10061 enc
29	412	9.9	75	23	AA145349 Human peptide enco
30	408.5	9.9	889	21	AA124216 Soluble neuropilli
31	401.5	9.7	538	22	AA102950 Angiotensin conver
32	401.5	9.7	600	22	AA102948 Angiotensin conver
33	401.5	9.7	644	20	AA106319 Human soluble neur
34	401.5	9.7	840	22	AA102949 Angiotensin conver
35	401.5	9.7	856	22	AA162478 Human NP-1 recepto
36	401.5	9.7	923	20	AA106317 Human neuropilin-1
37	401.5	9.7	923	20	AA123247 Human VEGF165R/NP-
38	401.5	9.7	923	20	AA196246 Human semaphorin r
39	401.5	9.7	923	23	AA122717 Human neuropilin-1
40	401.5	9.7	924	22	AA162476 Human VEGF(165)R/N
41	399.5	9.6	923	20	AA14562 Human semaphorin r
42	398.5	9.6	909	20	AA196254 Mouse semaphorin r
43	398.5	9.6	926	20	AA196255 Human semaphorin r
44	398.5	9.6	926	23	AA122718 Human neuropilin-2
45	398.5	9.6	931	20	AA106318 Human neuropilin-2

ALIGNMENTS

RESULT 1
AA170539
ID AA170539 standard; Protein; 669 AA.
AC AA170539;
XX
XX
XX
04-JUL-2000 (first entry)
XX
XX
Human Factor 8 Homologue.
XX
XX
Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW cerebroprotective; therapeutic; coagulation related disorder;
KW haemophilia; stroke; screening.
XX
XX
Homo sapiens.
XX
XX
WO200012532-A1.
XX
XX
09-MAR-2000.

1816 GGAGGAATGTTGGTACACTTCATCAAGATCTACCTTTAAACCAAGAGAAAGAA 1875
|||||
521 GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGlyLysGlu 540
1876 GCAGGCTATGCAGACTAGATCCTTACAACTCACAGGCGAGAGTGTATCATGCCTAT 1935
|||||
541 AlaGlyThrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyr 560
1936 GCTGAACCACTCCCAATACGGGGCTGAGTATGCAACCCCAATCATCATGACATGTCA 1995
|||||
561 AlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetAspMetSer 580
1996 GGGACCCCAACACTACAGTGTGAGCTGAGCTCCATCCATCCACTTCAAGCTACGGGGAAC 2055
|||||
581 GlyHisProThrThrSerValGlyGlnProSerThrPheLysAlaThrGlyAsn 600
2056 CAACCTCCCCCACTAGTGGAACTTACAACTACACTTCTCCAGGACTCACAGCTGCTCC 2115
|||||
601 GlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSer 620
2116 TCAGCCAGGCCAGTATGATACCCCGAAAGCTGGGAGCCAGGTCTACCTGCCCCAGAC 2175
|||||
621 SerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAsp 640
2176 GAATGCTGTACCAAGTCCAGACAGACACACAAAGATATCAGGACGAGGAAGGATGGG 2235
|||||
641 GluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGly 660
2236 GAATGCTGTATTTTAAAGAAATCCTT 2262
|||||
661 GlucyAspValPheLysGluIleLeu 669

RESULT 2
AAG75450
ID AAG75450 standard; Protein; 583 AA.
XX
AC AAG75450;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6214.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US265524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
DR N-PSDB; AAH34855.
XX

Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
useful for preventing, diagnosing and/or treating colorectal cancers -
Claim 11; Page 7657-7660; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
cancer-associated nucleic acid molecules (N) and proteins (P), where
the proteins are collectively known as colon cancer antigens. The colon
cancer antigens have cytostatic activity and can be used in gene

CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and AAH37789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX

Sequence 583 AA:

Alignment Scores:
Pred. No.: 3.28e-249 Length: 583
Score: 2931.00 Matches: 551
Percent Similarity: 99.28% Conservative: 2
Best Local Similarity: 98.92% Mismatches: 4
Query Match: 70.75% Indels: 0
DB: 22 Gaps: 0

US-10-060-830-1 (1-2280) x AAG75450 (1-583)

QY 4 GCGCGCCCGCTGGCGCGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 63
|||||
Db 1 AlaAlaProAlaTrpAlaAlaLeuProLeuSerArgSerLeuProCysSerAsnSer 20
QY 64 TCCTCCTTCTCCATGCCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 123
|||||
Db 21 SerSerPheSerMetProLeuPheLeuLeuLeuLeuValLeuLeuLeuLeuGlu 40
QY 124 GACGCTGGAGCCCAAGAGGTGATGGACACTGTACTAGGCGCTCTGAGAGTGA 183
|||||
Db 41 AspAlaGlyAlaGlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGly 60
QY 184 ACCCTTACATCCATAAATACCCACAGACCTATCCACAGACACTGTTGTGAATGGGAG 243
|||||
Db 61 ThrLeuThrSerIleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGlu 80
QY 244 ATCCCTGTAAAGATGGGAGAGAGCTTCGATCAAAATTTGGTACTTTGACATTCAGAT 303
|||||
Db 81 IleArgValLysMetGlyGluArgValArgIleLysPheGlyAspPheAspIleGluAsp 100
QY 304 TGTGATCTTGTCACTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 363
|||||
Db 101 SerAspSerCysHisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThr 120
QY 364 GAAATAGGCAATACTGTGGCTGGGTGGGAGTGGCAATGAACCATTCATTAATGAATCAAGGC 423
|||||
Db 121 GluIleGlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGly 140
QY 424 AATGAATACATCTGCTTCATGAGTGAATCCATGTTTCTGGACGGGATTTTGGCC 483
|||||
Db 141 AsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAla 160
QY 484 TCATACTCTGTTATAGATAAACAAGATCTAATTAATTAATTAATTAATTAATTAATTAAT 543
|||||
Db 161 SerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPhe 180
QY 544 TTGGAACCTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTCTCTCTCTCTCTCTCTCT 603
|||||
Db 181 LeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIle 200
QY 604 TCTGGAACAATTCCTCATGGATATAGATTCCTCCGCAATTCGATCGATCGCTGGTGTGAT 663
|||||
Db 201 SerGlyThrIleProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHis 220
QY 664 GCAGGAGTAGTGTCAACACACGTTGGCGCGCCCAATCACTGTTGTAAATAGTAAGGTATT 723

Db	221	AlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIle	240
Qy	724	CCCTATTATGAAGTCTTTGGCTAAACAGCTCACATCTCTGGTGGACACTTATCTACA	783
Db	241	ProTyrTyrGluSerSerLeuAlaAsnValThrSerValValGlyHisLeuSerThr	260
Qy	784	AGTCTTTTACATTAAGACAAGTGGATGTATATGACACATGGGATGGAGTCTGGTGTG	843
Db	261	SerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyVal	280
Qy	844	ATCGCGATCTCTAAATAACAGCATCTCTGTGTGAGTGGACTGACACACAGGCGAA	903
Db	281	IleAlaAspProGlnIleThrAlaSerSerValLeuGluThrPheAspHisThrGlyGln	300
Qy	904	GAGAACAGTGTGAAACCCAAAGCCAGGCTGAAACAACTGACCGCTTGGGCTGCT	963
Db	301	GluAsnSerTyrLysProLysLysAlaArgLeuLysLysProGlyProTyrAlaAla	320
Qy	964	TTTGCCACTGATGATACCATCGGTGTACAAATAGATTGAATAGGAAAGAAATRACA	1023
Db	321	PheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThr	340
Qy	1024	GCAATTAACACATGGATCCACCATGTGGAGCACAAATTAATCTGTCTGCCTACAGA	1083
Db	341	GlyIleIleThrThrGlyIleThrMetValGluHisAsnTyrTyrValSerAlaTyrArg	360
Qy	1084	ATCCCTGTACAGTGATGGGGCAGAAATGGACTGTGTACAGAGACCGTGGTGGAGCAA	1143
Db	361	IleLeuTyrSerAspAspGlyGlnTyrThrValTyrArgGluProGlyValGluGln	380
Qy	1144	GATAAGATATTTCAAGGAACAAAGATTATCACAGGATGTCGTAATACTTTTGGCA	1203
Db	381	AspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspValArgAsn***PheLeuPro	400
Qy	1204	CCAATTATTGACGTTTATTATAGTGAATCCTACCCATGCGACAGAAATTCCTATG	1263
Db	401	ProIleAlaArgPheIleArgValAsnProThrGlnTrpGlnLysIleAlaMet	420
Qy	1264	AAATGGAGTCTCGGATGTAGTTTATCTAAAGGTCGCTCCCAAACTTACTCAA	1323
Db	421	LysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGln	440
Qy	1324	CTCCACCTCTCGGACACGATGACCTCAAAACACTACAGCCCTCCAAAATAGCC	1383
Db	441	ProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAla	460
Qy	1384	AAAGTCGTGCCCCAAATTTACGCAACACATACACCTCGCAGTAGCATGATTCCT	1443
Db	461	LysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPhePro	480
Qy	1444	GCACAGACAGACAAACACTGCGCTGATATCAGAAATACTACCGTAACCTCCAAAT	1503
Db	481	AlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsn	500
Qy	1504	GTAAACAAAGATGTAGCGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1563
Db	501	ValThrLysAspValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThr	520
Qy	1564	ACTCTCATCTCATATTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1623
Db	521	ThrLeuIleLeuIleLeuValCysAlaThrPheIleArgAsnArgLysLysLysThrGlu	540
Qy	1624	GGACCTATGACTTACTTACTGGACCGGCGAGTGTGGTGGAAAGCAAGT	1674
Db	541	GlyThrTyrAspLeuProTyrTyrAspArgAlaGlyAsnSerArgGlyLeu	557
RESULT 3			
AAE22716			
ID	AAE22716 standard; Protein: 385 AA.		
XX			
AC			
XX	AAE22716;		

09-AUG-2002 (first entry)

Human neuropilin-Hy2 protein.

Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation; neurodegenerative disease; Alzheimer's disease; learning; angiogenesis; thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory; ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide; wound healing; tissue repair; Parkinson's disease; Huntington's disease; anyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord; cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis; human immunodeficiency virus; HIV; autoimmune disorder; dermatological; systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer; autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease; myasthenia gravis; autoimmune inflammatory eye disease; gene therapy; neotrophic; neuroprotective; vulnarary; anticonvulsant; antiparasitic; cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic; immunosuppressive; chromosome 6q21.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..20

Protein /label= Signal_peptide

61..385

/note= "Human mature neuropilin-Hy2 protein"

WO200222815-Al.

21-MAR-2002.

12-SEP-2001; 2001WO-US28488.

11-SEP-2000; 2000US-0659671.

06-SEP-2001; 2001US-317902P.

(HYSE-) HYSEQ INC.

Tang YT;

WPI; 2002-393966/42.

N-PSDB; AAD35994.

Novel isolated human Neuropilin-Hyl and Neuropilin-Hy2 polypeptides useful for treating neurodegenerative diseases e.g. Alzheimer's disease, and for diagnosing and mapping genetic neuronal defects

Claim 3; Page 128-130; 152pp; English.

The invention relates to neuropilin-like polypeptide (neuropilin-Hyl and neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-like polypeptides and polynucleotides are useful in modulating neuronal growth regenerative capacity, treating neurodegenerative diseases, diagnosing and mapping genetic neuronal defects and degenerative diseases like Alzheimer's disease and for treating learning and memory disorders. They are also useful for inducing angiogenesis, neovascularisation, as well as organ growth and development e.g. heart and other tissues. Antagonists of neuropilin-like polypeptides are useful for treating cancers and other malignant diseases. Neuropilin is used to treat platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal nocturnal haemoglobinuria and is used in nerve tissue growth or regeneration, in wound healing, tissue repair and replacement and in sequences of the invention are useful for treating diseases of peripheral nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager syndrome, anyotrophic lateral sclerosis, and traumatic disorders e.g. spinal cord disorders, head trauma and cerebrovascular diseases e.g. stroke, ulcers, immune deficiencies and immune disorders, infections by hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses, mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g. multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes mellitus, graft-versus-host disease, myasthenia gravis and autoimmune

Human TANGO 229 polypeptide.

Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung; skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte; bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer; anapylaxis; hepatitis; multiple sclerosis; coronary artery disease; malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis; attention deficit disorder; Crohn's disease; gastroenteritis; goitre; hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism; muscular dystrophy; immuno-competence; vertebrae; blood; serum.

OS Homo sapiens.

XX Key Location/Qualifiers
 XX Peptide 1..34
 FT /note= "Signal peptide"
 FT Domain 35..455
 FT /note= "Extracellular domain"
 FT Protein 35..715
 FT /note= "Mature human TANGO 229"
 FT Domain 456..480
 FT /note= "Transmembrane domain"
 FT Domain 481..715
 FT /note= "Cytoplasmic domain"

XX W0200129088-A1.

XX 26-APR-2001.

XX 23-JUN-2000; 2000WO-US17386.

XX 19-OCT-1999; 99US-0420707.

XX (MILL-) MILLENNIUM PHARM INC.

XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

XX WPI; 2001-308477/32.

XX N-PSDB; AAS00660.

XX New isolated nucleic acid molecule for diagnosis, prevention, and therapy of human and other animal disorder, or as modulating agent for regulating cellular processes

XX Claim 8; Fig 1; 263pp; English.

XX The sequence represents human TANGO 229 polypeptide. This protein and similar others exhibit the ability to affect growth, proliferation, survival, differentiation, activity, morphology, or movement/migration of, e.g. T cells and cells of the heart, liver, pancreas, placenta, brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus tissue. They can be used as modulating agents for regulating cellular processes, thus, the proteins and their associated nucleic acids can be used to prognosticate, prevent, diagnose, or treat disorders associated with physiological processes. These disorders include abnormal blood coagulation, asthma, anapylaxis, hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria, atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention deficit disorder, Crohn's disease, gastroenteritis, goitre, hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and muscular dystrophy. Antibodies to disorders such as these can be made by providing a polypeptide of the invention to an immuno-competent vertebrate and harvesting blood or serum from the vertebrate.

XX Sequence 715 AA;

Alignment Scores:

Pred. No.: 7,46e-83 Length: 715
 Score: 1042.00 Matches: 264
 Percent Similarity: 52.14% Conservative: 113
 Best Local Similarity: 36.51% Mismatches: 238
 Query Match: 25.15% Indels: 108
 DB: 22 Gaps: 24

US-10-060-830-1 (1-2280) x AAU00670 (1-715)
 QY 88 CTCTGCTCTTACTTGTCTGCTCTGCTGAGGACGCTGGAGCCAGCAAGGTGAT 147
 DB 20 LeuAlaLeuLeuLeuAlaValSerAlaProLeuArgLeuGlnAlaGluGluLeuGlyAsp 39
 QY 148 GGATGTGGACACTGTACTAGGCCCCGTGAGAGTGAACCTTACATCAATACTACCCA 207
 DB 40 GlyCysGlyHisLeuValThrTyrglnAspSerGlyThrMetThrSerLysAsnTyrPro 59
 QY 208 CAGACCTATCCACACACTGTTGTGAATGGGAGATCCGTGAAGATGGGAGAGAGA 267
 DB 60 GlyThrTyrglnProAsnHisThrValCysGluLysThrIleThrValProLysGlyLysArg 79
 QY 268 GTTCGCATCAAAATTTGGTGACTTTGACATTTGACATTTCTGATTTCTTCTCACTTAATAC 327
 DB 80 LeuIleLeuArgLeuGlyAspLeuAspIleGlu--SerGlnThrCysAlaSerAspTyr 98
 QY 328 TTGAGAATTTATAATGAATTTGGAGTCAAGCAAACTGAAATAGGCAAACTACTGTGGCTG 387
 DB 99 LeuLeuPheThrSer-----SerSerAspGlnTyrglyProTyrCysGly--- 113
 QY 388 GGGTTGCAATGAACCATTTCAATGAATCAAAAGGCAATGAATCAATCAATCTGCTTTCATG 447
 DB 114 SerMetThrValProLysGluLeuLeuLeuAsnThrSerGluValThrValArgPheGlu 133
 QY 448 AGTGAATCCATGTTCTGGACGGGATTTTGGCCCTCATCTCTGTATAGATAAACA 507
 DB 134 SerGlySerHisIleSerGlyArgGlyPheLeuLeuThrTyrglnAlaSerSerAspHisPro 153
 QY 508 GATCTAATTAATCTGTTGGACACTGCATCCAAATTTTGGAACTGAGTTCAGTAAGTAC 567
 DB 154 AspLeuIleThrCysLeuGluArgAlaSerHisTyrLeuLysThrGluTyrSerLysPhe 173
 QY 568 TGCCACGCTGTTGCTGCTCTCTTTTGTGAGATATCTGGAACAATCTCCTCATGGATAT 627
 DB 174 CysProAlaGlyCysArgAspValAlaGlyAspIleSerGlyAsnMetValAspGlyTyr 193
 QY 628 AGACATCTCTGCCATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 687
 DB 194 ArgAspThrSerLeuLeuCysLysAlaAlaIleHisAlaGlyIleIleAlaAspGluLeu 213
 QY 688 GCGGCCCCAAATCAGTGTGTAAATAGTAAGGTATTCCTTATTAAGAAAGTTCCTTGGCT 747
 DB 214 GlyGlyGlnIleSerValLeuGlnArgLysGlyIleSerArgTyrGluGlyIleLeuAla 233
 QY 748 AACACGTCACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 807
 DB 234 AsnGlyValLeuSerArgAspGlySerLeuSerAspLysArgPheLeuPheThrSerAsn 253
 QY 808 GGATGTTATGAACACTGGGATGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 867
 DB 254 GlyCysSerArgSerLeuSerPheGlu-----ProAspGlyGlnIleArgAla 269
 QY 868 TCATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 927
 DB 270 SerSerSerTrpGlnSerValAsnGlnSerGlyAspGlnValHisTrpSerProGlyGln 289
 QY 928 CCGAGGCTGAAACACCTGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 978
 DB 290 AlaArgLeuGlnAspGlnGlyProSerTrpAlaSerGlyAspSerSerAsnHisLys 309
 QY 979 --TACCAGTGTACAAATAGATTTGAATGAAGGAAAGAAATAACAGCATTAATAC 1035
 DB 310 ProArgGluTrpLeuGluIleAspLeuGlyGlyLysLysIleThrGlyIleArgThr 329
 QY 1036 ACTGGATCCACCATGGTGGAGCACAATTAATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1095
 DB 330 ThrGlySerThrGlnSerAsnPheAsnPheThrValLysSerPheValMetAsnPheLys 349
 QY 1096 GATGATGGCGAAGAAATGGACTGTGTACAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1155

```
Db 350 AsnAsnAsnSerLysTrpLysThrTyrLysGlyLeuValAsnAsnGluGluLysValPhe 369
Qy 1156 CAGGAACAAAGATTATCACCGAGGTGCGGTAATAACTTTTGGCACCATAATTATGCA 1215
Db 370 GlnGlyAsnSerAsnPheArgProValGlnAsnAsnPheIleProPheValAla 389
Qy 1216 CGTTTTATTAGATGATCCATCCCAATGCGACAGAAAATGCCATGAAAATGGAGCTG 1275
Db 390 ArgTyrValArgValProGlnThrTrpHisGlnArgIleAlaLeuLysValGluLeu 409
Qy 1276 CTCGGATGTCAGTTTATTCTTAAGGTGCTCCCAAACTTACTCAACCTCCACCTCCT 1335
Db 410 IleGlyCysGln-----IleThrGln----- 416
Qy 1336 CGGAACAGCAATGAC-----CTCAAAACACTACAGCCCTCCAAAATAGCC 1383
Db 417 -----GlyAsnAspSerLeuValTrpArgLysThrSerGlnSerThrSerValSerThr 434
Qy 1384 AAGGTGCTGCCCAAAATTTACGCAACCACTACAACCTCGCAGTAGCAATGAATTCCT 1443
Db 435 LysLysGluAspGluThrIleThrArgProIle----- 445
Qy 1444 GCACAGACAGAAACAACTGCCAGTCCATATACAGAAATACATCCGTAACCTCCAAT 1503
Db 446 ---ProSerGluGluThr-----SerThrGlyIleAsnIleThr----- 458
Qy 1504 GTAACCAAGATGTAGCGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1563
Db 459 -----ValAlaIleProLeuValLeuValLeuValValLeuValPheAlaGlyMet 474
Qy 1564 ACTCTATTCTCATATTAGTGTGTGCTGGCACTGGAGAAAGAAAACAACTGAA 1623
Db 475 GlyIlePheAlaAlaPhe-----ArgLysLysLysLysLys 486
Qy 1624 GGCACCTATGACTTACTTACTG-----GACCGCGCAGGTTGGTGGAAAGGA 1671
Db 487 GlySer-----ProTyrGlySerAlaGluAlaGlnLysThrAspCysTyrLysGln 503
Qy 1672 ATGAGCAGTTTCTCTCTGCAAGACGAGTGACCATGAGGAACCCCA---GTTCGGTAT 1728
Db 504 IleLysTyr-----ProPheAlaArgHis 511
Qy 1729 AGCAGACGCGAA-----GTTAATCAGCTGAGTCCAGGAAGTCAC---ACA 1773
Db 512 GlnSerAlaGluPheThrIleSerTyrAspAsnGluLysGluMetThrGlnLysLeuAsp 531
Qy 1774 GTGCTGAGGCTGACTGTCAGATATGCTCAGCACTGCTAGGAGAAATGTGTGGTACA 1833
Db 532 LeuIleThrSerAspMetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrVal 551
Qy 1834 CTTTCATCAAGATCTACTTTAAACCA-----GAAGAAGAAAGAGCAGGCTATGCA 1887
Db 552 ThrArgLysGlySerThrPheArgProMetAspThrAspAlaGluAlaGly---Val 570
Qy 1888 GACCTAGATCCT-----TACAACCTCAGCAGGAGGAAGTTTATCATCGCTATGCT 1938
Db 571 SerThrAspAlaGlyGlyHisTyrAspCysProGlnArgAlaGlyArgHisGluTyrAla 590
Qy 1939 GAACCACTCCCAATACGGGCGCTGAGTATGCAACCCCAATCATCATGACATGTCAGGG 1998
Db 591 LeuProLeuAlaProGluProGluTyrAlaThrProIleVal-----GluArg 607
Qy 1999 CACCCCACTTCACTGTTGTCACCCCTCCATCCACTTCAAGGCTACGGGGAACCAA 2058
Db 608 HisValLeuArgAlaHisThrPheSerAlaGlnSerGlyTyrArgValProGlyProGln 627
Qy 2059 CCT-----CCCCCACTA-----GTGGGAAT 2079
Db 628 ProGlyHisLysHisSerLeuSerSerGlyPheSerProValAlaGlyValGlyAla 647
Qy 2080 TACAATACACTTCTCTCCAGGATGACAGTGTCTCTCCAGCCAGCCAGCCAGTATGATACC 2139
Db 648 GlnAspGlyAspTyrGlnArgProHisSerAlaGlnProAlaAspArgGlyTyrAspArg 667
```

```
Qy 2140 CCGAAGCT 2148
Db 668 ProLysAla 670
```

RESULT 8

ABG01298

ID ABG01298 standard; Protein; 197 AA.

XX

AC ABG01298;

XX

DT 13-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #1289.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WC200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

XX

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

XX

DR N-PSDB; AAS65485.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX

PS Claim 20; SEQ ID NO 31657; 103pp; English.

XX

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 197 AA;

XX

Alignment Scores:

Pred. No.:	8.74e-83	Length:	197
Score:	1039.00	Matches:	197
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	25.08%	Indels:	0

DB: 22 Gaps: 0

US-10-060-830-1 (1-2280) x ABG01298 (1-197)

QY 1672 ATGAAGCAGTTTCTCTCTGCAAAAGCAGTGGACCATGAGGAACCCAGATTCGCTATAGC 1731

Db 1 MetLysGlnPheLeuProAlaLysLeuValAlaValAspHisGluGluThrProValArgTyrSer 20

QY 1732 AGCAGCGAAGTTAATCACTGAGTCCAGAGAGAGTCCACAGCTGCTGACGCTGACTCT 1791

Db 21 SerSerGluValAsnHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSer 40

QY 1792 GCAGAGTATGCTACGCCACTGCTGAGGAGAAATCTGTTGGTACACTTCATCAAGATCTACC 1851

Db 41 AlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThr 60

QY 1852 TTTAAACAG 1911

Db 61 PheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerPro 80

QY 1912 GGGCAGGAAGTTTATCATGCTTATGCTGAACCACTCCCAATACGGGCGCTGAGTATGCA 1971

Db 81 GlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAla 100

QY 1972 ACCCAATATCATGAGCATCTCAGGGCACCCCAACTTCAGTTGGTCAGCCCTCCACA 2031

Db 101 ThrProIleIleMetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThr 120

QY 2032 TCCACTTTCAGGCTACGGGAGAACCACTCCCACTAGTGGGAACTTACAAATACACTT 2091

Db 121 SerThrPheLysAlaThrGlyAsnGlnProProProLeuValGlyThrTyrAsnThrLeu 140

QY 2092 CTCCTCCAGGACTGACAGCTGCTCTCAGCCAGGCGCCAGTATGATACCCGAAAGCTGGG 2151

Db 141 LeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGly 160

QY 2152 AAGCAGGCTACCTGCCCCAGACGAATGGTGTACCAAGTGGTACCAAGTGGTACCAAG 2211

Db 161 LysProGlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGlu 180

QY 2212 GTATCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2262

Db 181 ValSerGlyAlaGlyArgAspGlyGluCysAspValPheLysGluLeuLeu 197

RESULT 9

AAU00629

ID AAU00629 standard; Protein: 586 AA.

XX AC AAU00629;

XX 29-AUG-2001 (first entry)

DE Novel human protein (NHP) sequence #2.

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

KW obesity; high blood pressure; connective tissue disorder; infertility;

KW NHP-mediated pathway.

XX Homo sapiens.

XX WO200129219-A1.

PN 26-APR-2001.

XX 08-OCT-2000; 2000WO-US28798.

XX 19-OCT-1999; 99US-0160285.

PR 18-FEB-2000; 2000US-0183583.

XX (LEXI-) LEXICON GENETICS INC.

PA Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

XX

DR WPI: 2001-290917/30.

DR N-PSDB: AAS00614.

XX Novel nucleic acid encoding human CUB-domain containing protein, useful for drug screening, diagnosis and treatment of physiological disorders or diseases.

PT Claim 2; Page 27-28; 33pp; English.

PS The sequence represents a novel human protein (NHP) containing a CUB domain (an extracellular domain). CUB proteins have been associated with regulating development, modulating cellular processes and preventing infectious disease. NHP nucleotide sequences are useful for gene therapy of physiological disorders or diseases. NHP oligonucleotides are useful as hybridisation probes for screening libraries and assessing gene patterns. NHP nucleotide sequences are useful for detecting mutant or inappropriately expressed NHPs (for example, those proteins associated with obesity, high blood pressure, connective tissue disorders and infertility) for the diagnosis of a disease. The polynucleotides may also be used in screening for drugs effective in the treatment of symptomatic or phenotypic manifestations of perturbing the normal function of NHP in the body. Nucleotide constructs encoding NHP products are used to genetically engineer host cells to express such products in vivo. These host cells allow for the identification of compounds that bind to NHP receptors or trigger NHP-mediated pathways.

XX Sequence 586 AA;

Alignment Scores:

Pred. No.: 1,14e-72 Length: 586

Score: 926.00 Matches: 225

Percent Similarity: 53.75% Conservative: 105

Best Local Similarity: 36.64% Mismatches: 198

Query Match: 22.35% Indels: 86

DB: Gaps: 18

US-10-060-830-1 (1-2280) x AAU00629 (1-586)

QY 7 GCCCCGCTGGCGCGCTCCCTCTCCGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 66

Db 43 AlaProSerGlyValMetValProGlyAlaArgGlyGlyAlaLeuAlaAlaAla 62

QY 67 TCTTCTCCAGCCCTCTGCT 126

Db 63 GlyArgGlyLeu-----LeuAlaLeuLeuLeuAlaValSerAlaProLeuArgLeu 79

QY 127 GCTGGAGCCAGCAAGGTGATGGATGGACACTGACTAGGCCCTGAGAGTGAAC 186

Db 80 GlnAlaGluGluLeuGlyAspGlyCysGlyHisLeuValThrTyrGlnAspSerGlyThr 99

QY 187 CTTACATCCATAAATACTACCCACAGACCTATCCCAACAGCACTGTTGTGAATGGAGATC 246

Db 100 MetThrSerLysAsnTyrProGlyThrTyrProAsnHisThrValCysGluLysThrIle 119

QY 247 COTGTAAGATGGGAGAGAGAGTTCGCATCAAAATTTGGTGTGACTTTGACATTCGAATTC 306

Db 120 ThrValProLysGlyLysArgLeuLeuLeuArgLeuGlyAspLeuAspIleGlu---Ser 138

QY 307 GATTCTGTGCACTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 366

Db 139 GlnThrCysAlaSerAspTyrLeuLeuPheThrSer-----SerSerAspGln 154

QY 367 ATAGGCAATACTGTGCTGGGTTCGAAATGAACCATTCATTAATGAATCAAAAGGCAAT 426

Db 155 TyrGlyProTyrCysGly---SerMetThrValProLysGluLeuLeuAsnThrSer 173

QY 427 GAAATCACATTCGTGTTATGAGTGAATCCATGTTTCTGGAGCGCGGATTTTGGCCATCA 486

Db 174 GluValThrValArgPheGluSerGlySerHisIleSerGlyArgGlyPheLeuLeuThr 193

QY 487 TACTCTGTATAGATAAACAAGATCTAATTAATTAATTAATTAATTAATTAATTAATTAAT 546

Db 194 TyrAlaSerSerAspHisProAspLeuIleThrCysLeuGluAlaSerHisThrLeu 213

Alignment Scores:

Pred. No.: 1.5e-72 Length: 539
 Score: 924.50 Matches: 221
 Percent Similarity: 54.34% Conservative: 98
 Best Local Similarity: 37.65% Mismatches: 185
 Query Match: 22.31% Indels: 63
 DB: 22 Gaps: 17

US-10-060-830-1 (1-2280) x AAU00630 (1-539)

```

QY 88 CTCCTGCTTACTTGTCTGCTGCTGCTGAGGAGCCCTGGAGCCAGCAAGGTGAT 147
DB 20 LeuAlaLeuLeuAlaValSerAlaProLeuArgLeuGlnAlaGluLeuGlyAsp 39
QY 148 GGATGGACACACGTACTAGCCCTGAGAGTGAACCTTACATCCATAACACCA 207
DB 40 GlyCysGlyHisLeuValThrTyrGlnAspSerGlyThrMetThrSerLysAsnTyrPro 59
QY 208 CAGACCTATCCCAACAGCACCTTTGTGAATGGGAGATCCGTGTAAGATGGGAGAGA 267
DB 60 GlyThrTyrProAsnHisThrValCysGluLysThrIleThrValProLysGlyLysArg 79
QY 268 GTTCCATCAATTTGGTGACTTTGACATTTGAATCTGATCTGTGCTACATTTAATTAC 327
DB 80 LeuIleLeuArgLeuGlyAspLeuAspIleGlu---SerGlnThrCysAlaSerAspTyr 98
QY 328 TTGGAATTTAATGAATTGGAGTCAGCAGCAACTGAAATAGGCAAAATACTGTGCTCG 387
DB 99 LeuLeuPheThrSer-----SerSerAspGlnTyrGlyProTyrCysGly--- 113
QY 388 GGGTTGCAATGAACCATTTCAATTGAATCAAAAGCAATGAATCAATGCTGTTTCATG 447
DB 114 SerMetThrValProLysGluLeuLeuLeuAsnThrSerGluValThrValArgPheGlu 133
QY 448 AGTGAATCCATGTTTCGGACCGGATTTTGGCCCTCATCTGCTGTTATAGATAACAA 507
DB 134 SerGlySerHisIleSerGlyArgGlyPheLeuLeuThrTyrAlaSerSerHisPro 153
QY 508 GATCAATTAATCTGTTGGACACTGCATCCAAATTTTGGAACTTCACTGAGTTCAGTAAGTAC 567
DB 154 AspLeuIleThrCysLeuGluArgAlaSerHisTyrLeuLysThrGluTyrSerLysPhe 173
QY 568 TCCCAAGCTGTGTGCTGCTCTTGTGCTGATATCTGGAACAATCTCTCATGGATAT 627
DB 174 CysProAlaGlyCysArgAspValAlaGlyAspIleSerGlyAsnMetValAspGlyTyr 193
QY 628 AGAGATTCTCCCATTTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 687
DB 194 ArgAspThrSerLeuLeuCysLysAlaAlaIleHisAlaGlyIleAlaAspGluLeu 213
QY 688 GCGGCGCAATCAGTGTCTTAATAGTAAAGTATTCCCTATTATGAAAGTCTTTTGGCT 747
DB 214 GlyGlyGlnIleSerValLeuGlnArgLysGlyIleSerArgTyrGluGlyIleLeuAla 233
QY 748 AACACGTCACATCTGTGCTGGGACACTTCTACAAAGTCTTTTACATTTTAAGACAAGT 807
DB 234 AsnGlyValLeuSerArgAspGlySerLeuSerAspLysArgPheLeuPheThrSerAsn 253
QY 808 GGATGTTATCGAACACTGGGATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 867
DB 254 GlyCysSerArgSerLeuSerPheGlu-----ProAspGlyGlnIleArgAla 269
QY 868 TCATCTGCTGCTGGAGTGGAGTACCACACAGGAGAGAACAGTGTGGAACCCCAAAAAA 927
DB 270 SerSerSerTyrGlnSerValAsnGluSerGlyAspGlnValHisTrpSerProGlyGln 289
QY 928 GCGAGCTGAAACAAACCTGGACCGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 978
DB 290 AlaArgLeuGlnAspGlnGlyProSerTyrAlaSerGlyAspSerSerAsnAsnHisLys 309
QY 979 ----TACCAGTGGTTACAAATAGATTGAATAGGAAAGAAATAACAGGCATTATAACC 1035
DB 310 ProArgGluTrpLeuGluIleAspLeuGlyGluLysLysIleThrGlyIleArgThr 329

```

```

QY 1036 ACTGGATCCACCATGGTGGAGCACAATTACTATGTGCTGCTACAGAAATCCTGTACAGT 1095
DB 330 ThrGlySerThrGlnSerAsnPheAsnPheTyrValLysSerPheValMetAsnPheLys 349
QY 1096 GATGATGGCAGAAATGGACTGTGTACAGAGCCCTGGTGTGGAGCAAGATAGATATTT 1155
DB 350 AsnAsnAsnSerLysTyrThrTyrLysGlyIleValAsnAsnGluGluLysValPhe 369
QY 1156 CAAGAAACAAAGATTATCACCAGGATGCGTAATAACTTTTTCGCCCAATATTATGCA 1215
DB 370 GlnGlyAsnSerAsnPheArgAspProValGlnAsnAsnPheIleProIleValAla 389
QY 1216 CTTTTATTAGAGTGAATCCTACCCAAATGGCAGCAGAAAATGCCATGAAAATGGAGCTG 1275
DB 390 ArgTyrValArgValProGlnThrTyrHisGlnArgIleAlaLeuLysValGluLeu 409
QY 1276 CTCGGATGTCATTTATCTCTAAAGTCTGCTCCAAACTTACTCAACTCCACTCTCT 1335
DB 410 IleGlyCysGln-----IleThrGln----- 416
QY 1336 CGGAACAGCAATGAC-----CTCAAAACACTACAGCCCTCCAAAATAGCC 1383
DB 417 -----GlyAsnAspSerLeuValTrpArgLysThrSerGlnSerThrSerValSerThr 434
QY 1384 AAAGTCTGTCGCCCAAAATTTACGCAACCACTCAACCTCGCAGTAGCAATGAATTCCT 1443
DB 435 LysLysGluAspGluThrIleThrArgProIle----- 445
QY 1444 GCACAGACAGAACAAACTGCCAGTCTGTATATCAGAAATACTACCTAATCCAAAT 1503
DB 446 ---ProSerGluGluThr-----SerThrGlyIleAsnIleThrThr----- 458
QY 1504 GTAACCAAGATCTAGCGTGGCTGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1563
DB 459 -----ValAlaIleProLeuValLeuLeuValValLeuValPheAlaGlyMet 474
QY 1564 ACTCTCATCTCATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1623
DB 475 GlyIlePheAlaAlaPhe-----ArgLysLysLysLysLys 486
QY 1624 GGCACCTATGACTTACTTACTG-----GACCGGCGAGCTGTTGGTGGAAAGGA 1671
DB 487 GlySer-----ProTyrGlySerAlaGluAlaGlnLysThrAspCysTrpLysGln 503
QY 1672 ATGAAGCAGTTTCTTCTCTCAAAAGCAGTGGACCATGAGGAAACCCCA---GTTCGCTAT 1728
DB 504 IleLysTyr-----ProPheAlaArgHis 511
QY 1729 AGCAGCAGCGAA-----GTTAATCACCTGAGTCCAAAGAGAAAGTACCACAGCTGTG 1782
DB 512 GlnSerAlaGluPheThrIleSerTyrAspAsnGluLysGluMetThrGlnLysLeuAsp 531
QY 1783 GGTGACTGTGAGAGTATGCT 1803
DB 532 LeuIleThrSerAspMetAla 538

```

RESULT 11

AAU00628 standard; Protein; 487 AA.

ID AAU00628;

AC AAU00628;

XX 29-AUG-2001 (first entry)

DT Novel human protein (NHP) sequence #1.

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

XX Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;

Db	384	ProLeuThrHisSerGluProGluTyrAlaThrProIleValGluArgHisLeuLeuArg	403
Qy	1984	-----ATGCACATGTCAGGGCCACCACCACAACTTCACGTTGGTCAGCCCTCCACA	2031
Db	404	AlaHisThrPheSerThrGlnSerGlyTyrArgValProGly---ProArgProThrHis	422
Qy	2032	TCACATTTCAAGGCTACGGGACACCACTCCCCACACTAGTGGGACACTTACAACTACACTT	2091
Db	423	GluHisSerHisSerSerGlyGlyPheProProAlaThrGlyAlaThrGlnValGluSer	442
Qy	2092	CTCTCAGGACTCACAGCTGCTCTCAGCCAGCCAGCCAGTATGATACCCGAAAGCTGGG	2151
Db	443	TyrGlnArgProAlaSerProLysProValGlyGlyTyrAspLysPro---AlaAla	461
Qy	2152	AAGCCAGGTTACTCTCCCCAGCAGCAATTTGGTTACCAGGTGCCAGACAGCACACAGAA	2211
Db	462	SerSerPheLeuAspSerArgAsp-----ProAlaSerGlnSerGln	475
Qy	2212	GTATCAGGAGCAGGAGGGATGGG	2235
Db	476	MetThrSerGlyGlyAspaspGly	483
RESULT 13			
ID	AAE22715		
XX	AAE22715 standard; Protein; 398 AA.		
AC	AAE22715;		
XX			
DT	09-AUG-2002 (first entry)		
XX			
XX	Human neuropilin-Hy1 protein.		
XX			
KW	Human, neuropilin-like polypeptide; neuropilin-Hy1, neovascularisation		
KW	neurodegenerative disease; Alzheimer's disease; learning; angiogenesis		
KW	thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory		
KW	ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;		
KW	wound healing; tissue repair; Parkinson's disease; Huntington's disease;		
KW	anyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cor		
KW	neurovascular; stroke; ulcer; immune; infection; multiple sclerosis;		
KW	human immunodeficiency virus; HIV; autoimmune disorder; dermatological		
KW	systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer		
KW	autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;		
KW	myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;		
KW	nootropic; neuroprotective; vulvarey; anticonvulsant; antiparasitic;		
KW	neuroprotective; tranquiliser; virucide; antibacterial; cytostatic;		
KW	immunosuppressive; chromosome 6q21.		

XX WO20022815-A1.

XX
PD 21-MAR-2002.

12-SEP-2001;

PR 11-SEP-2000; 2000US-0659671.
PR 05-SEP-2001; 2001US-2179025

XX
PA (HYSE-) HVSE0 INC

XX
PI
Tang YT:XX
DR WPI; 2002-393966/42.

DR N-PSDB; AAD33992
XX

PT useful for treating

XX
PC
Class 3: Page 123-124

XX The invention relates
CC

CC neuropilin-Hy2) and

OS Homo sapiens.
 PN WO200222780-A2.
 XX PD 21-MAR-2002.
 XX PF 11-SEP-2001; 2001WO-US28590.
 XX PR 11-SEP-2000; 2000US-0659671.
 PR 06-SEP-2001; 2001US-0659671.
 XX XX
 PA (TANG/) TANG T Y.
 XX XX
 PI Tang TY;
 XX WPI; 2002-351881/38.
 DR N-PSDB; ABK49565.
 XX
 PT New neuropilin-like polypeptides for diagnosing, preventing and
 PT treating neurological conditions and disorders, cancers, and for
 PT inducing angiogenesis and neovascularisation -
 XX
 PS Claim 3; Page 118-120; 144pp; English.
 XX
 CC The invention relates to an isolated neuropilin-like polypeptide
 CC including neuropilin-Hyl and neuropilin-Hy2, the full length cDNAs
 CC encoding the proteins and the coding regions of the cDNAs. Also included
 CC is a nucleic acid array comprising the cDNAs attached to a surface used
 CC for detecting full-matches or mismatches to the cDNAs. The genes
 CC for neuropilin-Hyl and Hy2 are located on human chromosome 6q21.
 CC The nucleic acid array is useful for detecting full-matches or mismatches
 CC to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful
 CC in modulating neuronal growth, regenerative capacity, treating
 CC neurodegenerative diseases, learning and memory disorders, diagnosing and
 CC mapping genetic neuronal defects and degenerative diseases like
 CC Alzheimer's disease, for inducing angiogenesis, and neovascularisation
 CC and organ growth and development (e.g. the heart). The nervous system
 CC disorders include lesions of central or peripheral nervous systems,
 CC including traumatic lesions, ischaemic lesions, infectious lesions,
 CC degenerative lesions, lesions associated with nutritional diseases or
 CC disorders, neurological lesions, and lesions caused by toxic substances.
 CC The neuropilin-like proteins and cDNAs are also useful as markers for
 CC cancers. The neuropilin-like proteins are useful for regulating cell
 CC proliferation, cell differentiation, stem cell growth factor activity,
 CC for inducing proliferation of neural cells, regeneration of nerve and
 CC brain tissue, for treatment of central and peripheral nervous system
 CC diseases, and neuropathies, such as Parkinson's disease,
 CC Huntington's disease, amyotrophic lateral sclerosis, to regulate
 CC haematopoiesis and treat myeloid and lymphoid cell disorders, various
 CC anaemias, and platelet disorders, such as thrombocytopaenia,
 CC regeneration and treatment of lung or liver fibrosis, reperfusion
 CC injury in various tissues and as a food supplement or molecular
 CC weight marker. The cDNAs are useful in gene identification, genome
 CC mapping, transgenics, as hybridisation probes, for primer design, for
 CC gene chips and as a DNA antigen. The present sequence represents
 CC neuropilin-Hyl.
 XX

SQ Sequence 398 AA;

Alignment Scores:

Pred. No.:	8,72e-39	Length:	398
Score:	540.90	Matches:	138
Percent Similarity:	45.84%	Conservative:	55
Best Local Similarity:	32.78%	Mismatches:	117
Query Match:	13.05%	Indels:	111
DB:	23	Gaps:	11

US-10-060-830-1 (1-2280) x AAU79459 (1-398)

QY 139 CAAGGTGATGGTGGACACACTGTCAGTACGGCCCTGAGAGTGGACCCCTTACATCCATA 198

Db 3 GluGlyAspGlyCysGlyHisLeuValThrTyrGlnAspSerGlyThrMetThrSerLys 22

QY 199 AACTACCCACAGACCTATCCCAACAGCAGCACTGTTTGTGAATGGGAGATCCGTGTAAGATG 258
 Db 23 AsnTyrProGlyThrTyrProAsnHisThrValCysGluLysThrIleThrValProLys 42
 QY 259 GGAGAGAGAGTTTCGATCAAAATTTGGTCACTTTCACATTTGAAGATTTCTGATTTCTGTGCAC 318
 Db 43 GlyLysArgLeuIleLeuArgLeuGlyAspLeuAspIleGlu---SerGlnThrCysAla 61
 QY 319 TTTAATTACTTGAAGAAATTAATAAGAAATTCAGAGTTCAGCAGAACTGAAATAGGCAATAC 378
 Db 62 SerAspTyrLeuLeuPheThrSer-----SerSerAspGlnTyrGlyProTyr 77
 QY 379 TGTGCTGTGGGTTGCCAAATCAATTCATTTGATTCATCAAAAGCAATGAATCAATCATG 438
 Db 78 CysGly---SerMetThrValProLysLeuLeuLeuLeuLeuLeuValThrVal 96
 QY 439 CTGTTTCATGAGTGAATCCATGTTTCTGGACGCGGATTTTGGCTCATCTCTCTTATA 498
 Db 97 ArgPheGluSerGlySerHisIleSerGlyArgGlyPheLeuLeuThrTyrAlaSerSer 116
 QY 499 GAT-----ThraspGlyThr-----AAACAAGATCTAAT 516
 Db 117 AspHisProGluSerGlnGlyAspArgProSerGluLysThrLeuAspGlnGlnSerArg 136
 QY 517 ACTTGTGTTGGACATGCATCCAAATTTTGGACCTGAGTTCAGTAAGTACTGCCCGCCT 576
 Db 137 ThrPheLeuAlaThrGlyThrThrPheValLysAspSerPheSer----- 151
 QY 577 GGTGTCTGCTCTCTTTGCTGAGATATCTGGAACAATTCCTCATGGATATAGAGATCC 636
 Db 152 -----ThraspGlyThr----- 155
 QY 637 TCGCATTTGTCATGGCTGGTGTGTCATGACAGAGTAGTGTCAACACAGTGGCGGCGCAA 696
 Db 156 SerLeuLeuCysLysAlaAlaIleHisAlaGlyIleIleAlaAspGluLeuGlyGln 175
 QY 697 ATCAGTGTGTAATAGTAAAGGTATCCCTATTATGAAGTCTTTGGCTTAAACACGTC 756
 Db 176 IleSerValLeuGlnArgLysGlyIleSerArgTyrGluGlyIleLeuAlaAsnGlyVal 195
 QY 757 ACATCT----- 762
 Db 196 LeuSerArgGluPheGluIlePheArgGluGlnLeuPheSerSerValLeuPheTyrSer 215
 QY 763 -----GTGFTGGGACAC 774
 Db 216 TrpGlyAsnThrValHisAlaValIleGluLeuMetPheProHisMetIleValTrpHis 235
 QY 774 ----- 774
 Db 236 SerGlyLysThrArgGluGlySerIleAlaGluGluGluGlyValProLysLeuTyr 255
 QY 775 -----TTATCTACAAGTCTTTTACATTTTAAAGACAAAGTGGATGT 813
 Db 256 LeuValIleGlnLysGlnGluLeuValGlnAspLeuValLeuValAlaThrValGlyCys 275
 QY 814 TATGGAACACTGGGGATGGAGTCTGCTGTGTGTCGGGATCCCTCAATATACAGCATCATCT 873
 Db 276 SerArgSerLeuSerPheGlu-----ProAspGlyGlnIleArgAlaSerSer 291
 QY 874 GTGCTGGAGTGGACTGCACACACAGCGGACAGACAGTTCGAAACCCCAAAAGCCAGG 933
 Db 292 SerTrpGlnSerValAsnGluSerGlyAspGlnValHisTrpSerProGlyGlnAlaArg 311
 QY 934 CTGAAAAAACCTGGACCGCGCTTGGGCTGCTTTTGCACCTGATGAA-----TAC 981
 Db 312 LeuGlnAspGlnGlyProSerTrpAlaSerGlyAspSerSerAsnAsnHisLysProArg 331
 QY 982 CAGTGGTTACAATAGATTGAATTAAGGAAAAAGAAATAACAGGCATTATACCTCGA 1041
 Db 332 GluTrpLeuGluIleAspLeuGlyGluLysLysIleThrGlyIleArgThrGly 351
 QY 1042 TCCACCATGGTGGACGACAAATTACTATGCTGCTGCTACAGATCTCTGTAGTATGAT 1101

QY 937 AAAAAACCTGGACCCCTTGGCTGCTTTTGGCACTGATGAATACCACTGCTTACAATA 996
Dbb 312 HisGlyAspAspAsnGlyTrpThrProAsnValAspSerAsnLysGluTyrLeuGlnVal 331
QY 997 GATTTGAATAAGGAAAGAAATACAGGCATTATTAACCACTGGA-----TCCACCATG 1050
Dbb 332 AspLeuArgPheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGlu 351
QY 1051 GTGGAGCAAAATTACTATGTCTGCTGCCTACAGAAATCCCTGTCACAGTATGATGGCAGAAA 1110
Dbb 352 ThrGlnLysGlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAsp 371
QY 1111 TCGACTGTGTACAGAGACCCCTGGTGTGGACGAAGATATATTCAGGAAACAAAGAT 1170
Dbb 372 TrpMetValTyrArgHisGly-----LysAsnHisLysValPheGlnAlaAsnAsp 389
QY 1171 TATCACCAGGATGTCGTAATAACTTTTGGCACAATATTGCACGTTTATTATAGATG 1230
Dbb 390 AlaThrGluLeuValLeuAsnLysLeuHisThrProLeuLeuThrArgPheIleArgIle 409
QY 1231 AATCCTACCCATGCGACAGAAATTCGCATGAAATGGAGCTGCTCGGATGTCAGTTT 1290
Dbb 410 ArgProGlnThrTrpHisLeuGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1291 ATTCTAAAGTCGCTCCCAAACTTAGTCAACCTCCACCTCCTCGGAACAGCAATGAC 1350
Dbb 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMet 440
QY 1351 CTCAAAAACACTACAGCCCTCCAAAATAGCAAAAGTCGTGCCCAAAATTT----- 1404
Dbb 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerThrArgGluTyrLeuTrp 460
QY 1405 -----ACGCAACCACTACACCTCGCAGTAGCAATGAATTCCTCGCACAGACAAA 1455
Dbb 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgAsnProGln 480
QY 1456 CAACAACCTGCCAGCTCTGATATACAGAAATACCTACCTGCTGCTGCTGCTGCTG 1515
Dbb 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----CGCTGGCTGAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1548
Dbb 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaMetGluAlaArgAlaPhe 520
QY 1549 GTCATGGTCTCCTACTCTCATTCCTCATATTAGTGTGCTGCTGCTGCTGCTGCTG 1599
Dbb 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1600 -----AGNAACAGAAAGAAAGAACTGAAGGCACC-----TATGACTTACCTTACTGG 1647
Dbb 541 ProArgThrGlnGlnProLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1648 GACCGGGCAGGTGGTGGAAAGGAATGAAGCAGTTCTTCCTGCAAAAGCAGTGGACCAT 1707
Dbb 559 -----AspIleArgArgPhe----- 563
QY 1708 GAGAAACCCAGTTCGCTATAGCAGCAGCAAGTATATCACTGAGTCCAGAGAGATC 1767
Dbb 564 GluProValProAlaGlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIle 583
QY 1768 ACCACAGTCTCGAGGCTGACTCTCAGAGTAT-----GCTCAGCCACTGGTAGGAGA 1821
Dbb 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1822 ATGTGTGTACTTCAATCAAGATCTACCTTTAAACCA-----GAAGAAGA 1869
Dbb 604 LeuGlyProThrLysSerGluGluThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAAGAGCAGGCTAT-----GCAGACCTAGATCCTTACACTCA 1908
Dbb 624 ThrGluCysGlyLeuAsnCysSerPheGluAspLysAspLeuGln-LeuProSerGln 643

QY 1909 CCAGGGCAGGAAGTTTATCATGCTATGCTGACCACTCCCAATTACGGGCGCTCAGTAT 1968
Dbb 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTrpMetTyrAspArgAlaLys 663
QY 1969 GCAACCCCAATCATCATGTCATGACATGTCCAGGCACCCCA----- 2005
Dbb 663 strPleuGlnSerThrTrpIleSerSerAlaAsnProAsnAspArgThrPheProAspAs 683
QY 2006 -----CAACTTCAG-----TTGCT 2019
Dbb 683 pLysAsnPheLeuLysLeuGlnSerAspGlyGlyArgGluGlyGlnPheGlyArgLeuIle 703
QY 2020 CAGCCCTCCACATCCACTTTCAAGGCTACGGGGAACCAACCTCCCCCACTAGTGGAACT 2079
Dbb 703 eserProPro-ValHisLeu-----ProArgSerProValCysMetGluP 718
QY 2080 TACAATACACTTCTCTC-----C 2097
Dbb 718 heGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuGlnValValArgGluAlaArg 738
QY 2098 AGGACTGACACTGCTC-----CTCAGCCAGGCCCTAGTATGATACCCGAAAGCTGGG 2151
Dbb 738 lngLysLeuLeuTrpValIle-ArgGluAspGlnGlySerGluTrpLysHisGly 757
QY 2152 AAGCCAGTCTACCTGCTCCCAAGCAATTTGTTACCAAGTCCAGGTCACAGACACACAGAA 2211
Dbb 758 ArgIleLeuProSerTyrAsp---MetGluTyrGlnIle-----ValPheGluGly 774
QY 2212 GTATCAGGAGCAGGAGGATGGGGAATGTGATGT 2247
Dbb 775 ValIleGlyLysGlyArgSerGlyGluIleSerIle 786

Search completed: January 21, 2003, 09:35:20
Job time : 73.2481 secs

Qy	88	CTCCTGCTCTACTTGC-----CTGCTCCTGCTGCTCGAGACGCTGA	132
Db		::: ::: :::	
Db	1	MetLeuLeuArgLeuLeuSerCysCysTrpLeuLeuCysSerLeuArgSerSerTrp	20
Qy	133	GCCACAGAGGTGATGGATGTCACACACTGACTAGGCCCTGAGAGTGGACCTTACA	192
Db		:::: ::: ::: ::: ::: ::: :::	
Db	21	AlaSerArgAsnAspLysCysGlyAspThrIleLysIleThrSerProSerTyrLeuThr	40
Qy	193	TCCATAACTACCCACAGACCTATCCACACAGCACTGTTTGTGAATGGAGATCCGTGTA	252
Db		:::: ::: ::: ::: :::	
Db	41	SerAlaGlyTyrProHisSerTyrProProSerGlnArgCysGluTrpLeuIleGlnAla	60
Qy	253	---AAGATGGAGAGAGAGTTCGCATCAAAATTTGGT---GACTTTACATTGAAGATTCT	306
Db		::: :::: ::: :::: ::: ::: ::: :::	
Db	61	ProGluHisTyrGlnArgIleMetIleAsnPheAsnProHisPheAspLeuGluAspArg	80
Qy	307	GATTCTTCTCACTTAATTAATTCTTGAGAAATATATAATGGAAATGGAGTCAGAGA	366
Db		::: ::: ::: ::: ::: ::: ::: :::	
Db	81	Glu---CysLysTyrAspTyrValGluValIleAspGlyAspAsnAlaAsnGlyGlnLeu	99
Qy	367	ATAGGCAAACTACTGTGTCTGGGTGCAAAATGAACCATCA---ATTGAATCAAAAGGC	423
Db		::: ::: ::: ::: ::: ::: ::: :::	
Db	100	LeuGlyLysTyrCysGly-----LysIleAlaProSerProLeuValSerThrGly	116
Qy	424	AATGAATCACAATGCTGTTTCATGAGTGGAAATCCATGTTTCTGGAGCCGGATTITGGCC	483
Db		:::: ::: ::: ::: :::	
Db	117	ProSerIlePheIleArgPheValSerAspTyrGluThrProGlyAlaGlyPheSerIle	136
Qy	484	TCATCACTCTGTTATAGATAAACAAGATCAATTACTTGTGTTGGACACTGCATCCA	543
Db			
Db	137	ArgTyrGluValPheLys-----	142
Qy	544	TTGGAACCTGAGTTCAGTACTAGTACCTGCCAGCTGGTTGCTGCTCCTTTGCTGAGATA	603
Db			
Db	143	ThrGlyProGlu-----CysSerArgAsnPheThrSerSer	154
Qy	604	TCTGGAACAATT-----CCTCATGGATATAGATATCC-----TCG	639
Db		:::	
Db	155	AsnGlyValIleLysSerProLysTyrProGluLysTyrProAsnAlaLeuGluCysThr	174
Qy	640	CCATTGTGCATGGCTGTGTCATGCAGGAGTAGTG-----	675
Db		:::	
Db	175	TyrIleIlePheAlaProLysMetGlnGluIleValLeuGluPheGluSerPheGluLeu	194
Qy	676	-----TCAACACGTTGGCGGCCCAATACGT-----GTTGTAATTACT	714
Db			
Db	195	GluAlaAspSerAsnAlaProGlyGlyGlnThrCysArgTyrAspTrpLeuGlyIleTrp	214
Qy	715	AAAGGTATCCC-----TATTATGAAAGTTCTTTGGCTAAC	750
Db			
Db	215	AspGlyPheProGlyValGlyProHisIleGlyArgTyrCysGlyGlnAsnThrProGly	234
Qy	751	AACGTCATCTGTGGTGGGACACTTATCTACAAAGCTCTTTTATACATTTAAGACAAGTGA	810
Db			
Db	235	ArgValArgSerPheThrGlyIleLeuSerMetIlePheHisThrAspSerAlaIleAla	254
Qy	810	-----	810
Db			
Db	255	LysGluGlyPhePheAlaAsnPheSerValValGlnSerAsnThrAspGluAspPheGln	274
Qy	811	TGTTATGAAACACTGGGATGGAGTCGTGGTGATCCGGGATCCTCAATAACAGCATCA	870
Db			
Db	275	CysLysGluAlaLeuGlyMetGluSerGlyGluIleHisPheAspGlnIleSerValSer	294
Qy	871	TCT-----GTGCTGGAGTGGACTGACCAACACAGGGCAA-----GAGACAGT	912
Db		:::: ::: ::: ::: ::: ::: ::: ::: :::	
Db	295	SerGlnTyrSerMetAsnTrpSerAlaGluArgSerArgLeuAsnTyrValGluAsnGly	314
Qy	913	TGGAAACCCAAAAGCCAGCTGAAAAAACCTTGGACCGCTTGGCGCTGCTTTGGCCACT	972
Db			
Db	315	TrpThrProGlyGluAspThrValLys-----	323

QY 973 GATGAATACCACTGGTTACAAATAGATTTGAATAAGGAAAGAAATAACAGGCATTATA 1032
Db 324 -----GluTrpIleGlnValAspLeuGluAsnLeuArgPheValSerGlyIleGly 340
QY 1033 ACCACTGGA-----TCCACCATGGGAGACAAATTAATCTATGCTGCTGCTACAGAATC 1086
Db 341 ThrGlnGlyAlaIleSerLysGluThrLysLysTyPheValLysSerTyLysVal 360
QY 1087 CTGTACACTGATGATGGGCAAGAAATGGACTGTGTACAGAGAGCCTGGTGGAGCAAGAT 1146
Db 361 AspIleSerSerAsnGlyAspTrpIleThrLeuLysAspGly-----AsnLysHis 378
QY 1147 AAGATATTCAAGGAAAGAAAGATATTCACCAGGATGTCGCTAATACITTTTGGCCACCA 1206
Db 379 LeuValPheThrGlyAsnThrAspAlaThrAspValValTyArgProPheSerLysPro 398
QY 1207 ATTATTGCACGTTTTATTAGAGTGAATCTACCCAATGGCAGCAGAAAATGGCATGAAA 1366
Db 399 ValIleThrArgPheValArgLeuArgProValThrTrpGluAsnGlyIleSerLeuArg 418
QY 1267 ATFGAGCTGCTCGGATGTGAGTTTATTCCTAAAGTCGTCCTCCAAACATTAATCTACACCT 1326
Db 419 PheGluLeuTyGlyCys-----LysIleThrAspTyr 429
QY 1327 CCACCTCCTCGGAACAGCAATGACCTC-----AAAAACACTACAGCC 1368
Db 430 ProCysSerArgMetLeuGlyMetValSerGlyLeuIleSerAspSerGlnIleThrAla 449
QY 1369 CTTCCAAAATAGCCAAAGTCGTGCGCCCAAAATTTACGCAACCACTACAACTCGCAGT 1428
Db 450 SerSerGlnValAspArgAsnTrpValProGluLeuAlaArgLeuValThrSerArgSer 469
QY 1429 -----AGCAATGAATTCCTCGCAGCAGACAGAA 1455
Db 470 GlyTrpAlaLeuProProSerAsnThrHisProTyThrLysGlu 484
RESULT 2
KFBO5
coagulation factor V precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
C:Accession: A42580; A36497
J. Guanto, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A:Title: The complete cDNA sequence of bovine coagulation factor V.
A:Reference number: A42580; MUID:92147638; PMID:1737753
A:Accession: A42580
A:Molecule type: mRNA
A:Residues: 1-2211 <GUI>
A:Cross-references: GB:M81440; NID:g163037; PIDN:AAA30512.1; PID:g163038
A>Note: sequence extracted from NCBI backbone (NCBI:80774, NCBI:P:80776)
R:Kalafatis, M.; Jenny, R.J.; Mann, K.G.
J. Biol. Chem. 265, 21580-21589, 1990
A:Title: Identification and characterization of a phospholipid-binding site
A:Reference number: A36497; MUID:91072354; PMID:2254316
A:Accession: A36497
A:Molecule type: protein
A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584; 1673-1676, 'X', 1678-1679, 'X'
R:Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.
Biochemistry 33, 13109-13116, 1994
A:Title: Determination of the disulfide bridges in factor Va heavy chain.

```

RESULT 2
KFBO5
coagulation factor v precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Nov-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
C:Accession: A42580; A36497
R:Guanto, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A>Title: The complete cDNA sequence of bovine coagulation factor V.
A:Reference number: A42580; MUID:92147638; PMID:1737753
A:Accession: A42580
A:Molecule type: mRNA
A:Residues: 1-2211 <GUI>
A:Cross-references: GB:M81440; NID:g163037; PIDN:AAA30512.1; PID:g163038
A>Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBI:P:80776)
R:Kalafatis, M.; Jenny, R.J.; Mann, K.G.
R. Biol. Chem. 265, 21580-21589, 1990
A>Title: Identification and characterization of a phospholipid-binding site
A:Reference number: A36497; MUID:91072354; PMID:2254316
A:Accession: A36497
A:Molecule type: protein
A:Residues: 1566-1570,'X', 1572-1581,'X', 1583-1584,1673-1676,'X',1678-1679,'X'
R:Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.
Biochemistry 33, 13109-13116, 1994
A>Title: Determination of the disulfide bridges in factor Va heavy chain.
A:Reference number: A35979; MUID:95034740; PMID:7947716
A:Contents: annotation
A>Note: 566-Cys and 617-Cys were shown to have free sulfhydryls
C:Comment: Factor V is activated by thrombin and partially by coagulation f
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the f
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; f
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid bind
F:1-28/pomatin: signal sequence #status predicted <SIG>
F:29-2211/Product: coagulation factor V #status predicted <NAT>
F:29-741/Product: coagulation factor Va heavy chain #status predicted <VAR>

```



```
F:1906-2061/Domain: discoidin I amino-terminal homology <DN1>
F:2065-2224/Domain: C2 <DC2>
F:2065-2224/Domain: discoidin I amino-terminal homology <DN2>
F:51_55_239_327_460_468_556_741_752_760_772_782_821_938_977_1074_1083_1103_1106_1479_
F:167_193_248_329_500_526_603_684_1725_1751_1907-2061_2066-2221/Disulfide bonds: #sta
F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:363_693_1546/Binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:382_1338/Binding site: carbohydrate (Asn) (covalent) #status absent
F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted
F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experiment
F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experime
F:1573-1574/Cleavage site: Arg-Ser (Chrombin) #status experimental

Alignment Scores:
Pred. No.:      5,41e-18      Length:      2224
Score:          351.00        Matches:     77
Percent Similarity: 59.04%    Conservative: 21
Best Local Similarity: 46.39% Mismatches:   58
Query Match:      8.47%      Indels:      10
DB:               1         Gaps:        3

US-10-060-830-1 (1-2280) x KFHU5 (1-2224)

QY 799 AAGACAAGTGATGTTATGAACACTGGGATGGAGTCGTGTGTATCGCGGATCCTCAA 858
   ::: :::::::::::::::::::: ||||| ::::: |||
Db 2062 GluValAsnGlycysSerThrProLeuGlyMetGluAsnGlyLysIleGluAsnLysGln 2081
   ::: :::::::::::::::::::: ||||| ::::: |||

QY 859 ATAACACATCATCTGTGCTGGAG-----TGGACTGCACCACACAGGGCAGAGACAGT 912
   ||||||| ::::: ||| :::::
Db 2082 IleThrAlaSerSerPheLysLysSerTrpTrpGlyAspTyr----- 2095
   ||||||| ::::: ||| :::::

QY 913 TGGAAACCCAAAAGCCAGCGCTGAAAAAACCTGGA-----CGCCTTGGCTGCTTTT 966
   ||||| ::::: ||||| ::::: |||
Db 2096 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2115
   ||||| ::::: ||||| ::::: |||

QY 967 GCCACTGATGAATACCACTGGTGTACAAATAGATTGAATAAGAAAGAAAATAACAGGC 1026
   ||||| ::::: ||||| ::::: |||
Db 2116 AlaAsnAsnLysGlnTrpLeuGluIleAspLeuLysIleLysLysIleThrAla 2135
   ||||| ::::: ||||| ::::: |||

QY 1027 ATTATACCACTGGATCCACCATGGTGGACACAAATTAATCTATGTCTGCCTACAGATC 1086
   ||||||| ::::: ||||| ::::: |||
Db 2136 IleIleThrGlnGlyCysLysSerLeuSerSerGluMetTyrValLysSerTyrThrile 2155
   ||||||| ::::: ||||| ::::: |||

QY 1087 CTGTACAGTAGTATGATGCCAGAAATGACTGTGTACAGAGAGCGCTGTGTGGAGCAAGAT 1146
   ||||||| ::::: ||||| ::::: |||
Db 2156 HistySerGluGlnGlyValGluTrpLysProTyrArgLeuLysSerSerMetValasp 2175
   ||||||| ::::: ||||| ::::: |||

QY 1147 AAGATATTTCAGGAAACAAAGATTATCACAGGATGTGCGTAAATAACTTTTTGCCACCA 1206
   ||||||| ::::: ||||| ::::: |||
Db 2176 LysIlePheGluGlyAsnThrAsnThrLysGlyHisValLysAsnPheAsnProprio 2195
   ||||||| ::::: ||||| ::::: |||

QY 1207 ATATTGACGTTTTATTAGTGAATCCATACCAATGGCAGCAGAAANATGCCATGAAA 1266
   ||||||| ::::: ||||| ::::: |||
Db 2196 IleIleSerArgPheIleArgValIleProLysThrTrpAsnGlnSerIleAlaLeuArg 2215
   ||||||| ::::: ||||| ::::: |||

QY 1267 ATGGAGCTGCTCGGATGT 1284
   ||||||| ::::: ||||| ::::: |||
Db 2216 LeuGluLeuPheGlyCys 2221
   ||||||| ::::: ||||| ::::: |||

RESULT 5
S65138
N:Alternate names: glycoprotein antigen MGp57/53, mammary gland - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C:Accession: S65138; G48394
R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245: 395-391, 1995
A>Title: Molecular cloning of glycoprotein antigens MGp57/53 recognized by monoclonal
A:Reference number: S65138; MUID:96125736; PMID:8541316
A:Accession: S65138
A>Status: preliminary; not compared with conceptual translation
```


Db 49 LeuArgPheGluLeuLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 68
QY 835 TCTGTGTGATCGCGGATCCTCAATAACAGCATCATCTGTCTGGAG---TGGACTGAC 891
Db 69 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerSerTyrlsThrTrpGlyLeu 88
QY 892 CACACAGGCAAGAACAGAGTGGTGAACCCCAAAAGCCAGGCTGAAACACCTGGA--- 948
Db 89 HisLeu-----PheSerTrpAsnProSerTyrlaAlaArgLeuAspLysGlnGlyAsn 105
QY 949 ---CGCCCTTGGCTGCTTGGCCACTGATGATACCATGCTGCTTCAATATGAT 1005
Db 106 PheAsnAlaTrpValAlaGlySerTyrlsGlyAsnAspGlnTrpLeuGlnValAspLeuGly 125
QY 1006 AAGGAAAGAAATAACAGGATTAATACCATGCTGATCCACCATGCTGAGCACAATTAC 1065
Db 126 SerSerLysGluValThrGlyIleThrGlnGlyAlaArgAsnPheGlySerValGln 145
QY 1066 TATGTCTGCTCAGCAATCCTGTACAGTGTATGATGGCAGAAATGGACTGTGTACAGA 1125
Db 146 PheValAlaSerTyrlsValAlaTyrlsSerAsnAspSerAlaAsnTrpThrGluTyrlsGln 165
QY 1126 GAGCTGTGTGGACACAGATAGATATTTCAAGGAACAAAGATATATCACCAGGATGTG 1185
Db 166 AspProArgThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisSerHisLys 185
QY 1186 CGTAATAACTTTTCCCAACCAATATTCACGCTTTTATTAGAGTGAATCCTACCCCAATGG 1245
Db 186 LysAsnLeuPheGluThrProIleLeuAlaArgTyrlsValArgIleLeuProValAlaTrp 205
QY 1246 CAGCAGAAATTCGCAATGAAATGGAGTGTGCTGCGATGT 1284
Db 206 HisAsnArgIleAlaLeuArgLeuGluLeuLeuGlyCys 218

RESULT 8

T11743
pP47 protein - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T11743
R:Enslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toepfe
Biol. Reprod. 58, 1057-1064, 1998
A:Title: Molecular cloning and characterization of p47, a novel boar sperm-associated z
A:Reference number: Z17325; MUID:98206817; PMID:9546740
A:Accession: T11743
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-409 <ENS>
A:Cross-references: EMBL:X11683; NID:g2652927; PIDN:CAA72379.1; PID:g2652928
A:Experimental source: testis
C:Function:
A:Description: may be involved in membrane remodeling and/or function as a zona pellucid
C:Superfamily: unassigned EGF-related proteins; EGF homology
F:6-40/Domain: EGF homology <EGF>

Alignment Scores:
Pred. No.: 2,71e-17 Length: 409
Score: 340.00 Matches: 76
Percent Similarity: 56.59% Conservative: 27
Best Local Similarity: 41.76% Mismatches: 68
Query Match: 8.21% Indels: 11
DB: 2 Gaps: 5

US-10-060-830-1 (1-2280) x T11743 (1-409)

QY 759 ATCTGT-----GGTGGGACATTAATCAAGTCTTTTACATTTAAGACAAAGTGATG 812
Db 233 IleCysHisArgGlyCysThrLeuArgPheGlu-LeuLeuGlyCysGluLeuSerGlyCys 252
QY 813 TTATGAACACTGGGATGAGTGTGTGTGATCGGGATCCCAATATACACATCATC 872
Db 252 sAlaGluProLeuGlyLeuLysAspAsnThrIleProAsnLysGlnIleThrAlaSer 272

QY 873 TGTGCTGGAG---TGGACTGACACACACAGG---CAAGAGAACACAGTTGGAACCCAAAA 926
Db 272 rPheTyrlsArgThrTrp-----GlyLeuSerAlaPheSerTrpTyrlsPropheTy 288
QY 927 ACCCAGGCTGAAAAAACCTGGACCG-----CCTTGGGCTGCTTTGGCCACTGATGAATA 980
Db 288 rAlaArgLeuAspAsnGlnGlyLysPheAsnAlaTrpThrAlaGlnSerAsnSerAlase 308
QY 981 CCAGTGGTTACAAATAGATTTTGAATAAGAAAGAAAAATAACAGCATTATTAACACCTGG 1040
Db 308 rGluTrpLeuGlnIleAspLeuGlySerGlnArgValThrGlyIleThrGlnGln 328
QY 1041 ATCCACCAGTGGTGGAGCACAATTAATGTCTGCTCCCTACAGAAATCCCTGTACAGTGTGA 1100
Db 328 yAlaArgAspPheGlyHisIleGlnTyrlsValAlaTyrlsValAlaTyrlsSerAspAs 348
QY 1101 TGGGCGAGAATCGACTGTGTACAGAGAGCCTGTGTGGAGCAAGATAGATATTTCAAGG 1160
Db 348 pGlyValSerTrpThrGluTyrlsArgAspGlnGlyAlaLeuGluGlyLysIlePheProGln 368
QY 1161 AAACAAGATTTATCACCAGGATGTGCTAATAACTTTTCCCAACCAATATTGCACTGT 1220
Db 368 yAsnLeuAspAsnAsnSerHisLysLysAsnMetPheGluThrProPheLeuThrArgPh 388
QY 1221 TATTAGATGAATCCTTACCATGGCAGCAGAAATTTGCCATGAAATGGAGCTGCTCGG 1280
Db 388 eValArgIleLeuProValAlaTrpHisAsnArgIleThrLeuArgValGluLeuLeuGln 408
QY 1281 ATGT 1284
Db 408 yCys 409

RESULT 9

JC4915
ags protein precursor - rat
N:Alternate names: O-acetyl-Gd3 ganglioside
C:Species: Rattus norvegicus (Norway rat)
C>Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
C:Accession: JC4915
R:Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A:Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
A:Reference number: JC4915; MUID:96374422; PMID:8780713
A:Accession: JC4915
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <OGU>
A:Cross-references: DDBJ:D84068; NID:g1620006; PIDN:BAAL2210.1; PID:g1620007
A:Experimental source: CST cell
C:Comment: This protein is required for the O-acetylation of disialoganglioside sial:
C:Genetics:
A:Gene: ags
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hor
F:1-21/Domain: signal sequence #status predicted <SIG>
F:28-60/Domain: EGF homology <EG1>
F:68-107/Domain: EGF homology <EG2>
F:110-267/Domain: discoidin I amino-terminal homology <DN1>
F:271-427/Domain: discoidin I amino-terminal homology <DN2>

Alignment Scores:
Pred. No.: 3.24e-17 Length: 427
Score: 339.00 Matches: 128
Percent Similarity: 38.66% Conservative: 56
Best Local Similarity: 26.89% Mismatches: 160
Query Match: 8.18% Indels: 132
DB: 2 Gaps: 19

US-10-060-830-1 (1-2280) x JC4915 (1-427)

QY 88 CTCTGCTCTTACTTGTCTCTGCTCTGCTGCTGAGACGCTGGAGCCCGACGAGTGTAT 147
Db 7 LeuAlaAlaLeuCysGlyValLeuLeuCysAlaSerGlyLeuPheAlaAlaSerGlyAsp 26

```
QY 148 GGATGT-----GGACACACTGTACTAGC----- 171
|||
Db 27 PheCysAspSerSerLeuCysLeuAsnGlyGlyThrCysLeuMetGlyGlnAspAsn 46
QY 172 -----CCTGAGAGTGAACCCCTTACATCCATAAATACCACAGACACTAT 216
|||
Db 47 IleTyrCysLeuCysProGluGlyPheThrGlyLeuValCysAsnGluThrGluGly 66
QY 217 CCCAACAGC---ACTGTTGTGAATGGGAGATCCGTTAAAGATGGGAGAGAGTTCGC 273
|||
Db 67 ProCysSerProAsnProCysPheHisAspAlaLysCysLeuValThrGluAspThr 85
QY 274 ATCAAAATTTGGTGACATTTGACATTGAAGATTCTGATTCTTCTGACCTTTAATTAATCTGAGA 333
|||
Db 86 ---GlnArgGlyAspIlePheThrGluTyrIleCysGlnCysProValGlyTyrSerGly 104
QY 334 ATTTATAATGGAAATGGATCGACAGAACTGAAATAGGC----- 372.
|||
Db 105 IleHisCysGluLeuGlyCysSer---ThrLysLeuGlyLeuGluGlyGlyAlaIleAla 123
QY 373 -----AAATACTGTGTT---CTGGGTTGCAATGAACCAT 405
|||
Db 124 AspSerGlnIleSerAlaSerSerValTyrMetGlyPheMetGlyLeuGln----- 140
QY 406 TCAATTAATGAATCAAAAGGCAATGAATGCATTCCTGCTTTCATGAGTGAATC----- 456
|||
Db 141 ---ArgTrpGlyProGluLeuAlaArgLeuTyrArgThrGlyIleValAsnAla 157
QY 456 ----- 456
Db 158 TrpThrAlaSerSerTyrAspSerLysProTrpIleGlnValAspPheLeuArgLysMet 177
QY 457 CAGTGTTCCTGGA----- 471
|||
Db 178 ArgValSerGlyValMetThrGlnGlyAlaSerArgAlaGlyArgAlaGluTyrLeuLys 197
QY 472 GGATTTTGGCCTCATPACTCTGTATAGATAAACAAGATCTAATTAATCTGTTGGACACT 531
|||
Db 198 ThrPheLysValAlaTyrSerLeuAspGlyArgArg----- 209
QY 532 GCATCCAAATTTTGGAACTGAGTCACTAAGTACTGCCACCTGCTGCTGCTGCTCTCT 591
|||
Db 210 ---PheGluPheIleGlnAspGlu----- 216
QY 592 TTGCTGAGATATCTGGAACAAATTCCTCAT-----GGATATAGAGATTCCTCGCCA 642
|||
Db 217 -----SerGlyThrGlyAspLysGluPheMetGlyAsnGlnAspAsnSer 232
QY 643 TTGTCATGGCTGCTGCTGATGAGAGTAGTGTCAAAACACCTGGCGGCCCAAAATCACT 702
|||
Db 233 Leu-----LysIleAsnMetPheAsnProThrLeuGluAlaGlnTyr--- 246
QY 703 GTTGTAAATAGTAAGGTATTCCTATATAGAAAGTCTTTGGCTAACACGTCACATCT 762
|||
Db 247 -----IleArgLeuTyrProValSerCysHisArgGlyCysThr--- 259
QY 763 GTGGTGGACACTATCTACAGTCTTTTACATTTAAGACAGTGGATGTTATGGAACA 822
|||
Db 260 -----LeuArgPheGluLeuLeuGlyCysGluLeuHisGlyCysSerGluPro 275
QY 823 CTGGGATGGAGTCTGCTGATCGGGGATCCTCAATAACACATCATCTGTGCTGGAG 882
|||
Db 276 LeuGlyLeuLysAsnThrIleProAspSerGlnIleThrAlaSerSer-----Ser 293
QY 883 TGGACTGACACACAGGCAAGAGACAGTGTGGAACCCCAAAAGCCAGGCTGAAAAA 942
|||
Db 294 TyrLysThrTrpAsnLeuArgAlaPheGlyTyrTrpTyrProHisLeuGlyArgLeuAspAsn 313
QY 943 CTGGGACCG-----CCTGGGCTGCTTTGGCACTGATGATACAGTGGTTACAATA 996
|||
Db 314 GlnGlyLysIleAsnAlaTrpThrAlaGlnSerAsnSerAlaLysGluTrpLeuGlnVal 333
```

```
QY 997 GATTGTAATAAGAAAGAAATAACAGCATTTATTAACCACTGGATPCCACCATGTGGAG 1056
|||
Db 334 AspLeuGlyThrGlnLysLysValThrGlyIleIleThrGlnGlyAlaArgAspPheGly 353
QY 1057 CACATTAATTAATGCTGCTGCCTACAGAAATCCTGTACAGTGTATGGGAGAAATGGACT 1116
|||
Db 354 HisIleGlnTyrValAlaSerTyrLysValAlaHisSerAspGlyValGlnTrpThr 373
QY 1117 GTGTACAGACGCCGTGCTGGAGCAAGATAAGATATTTCAAGAAACAAGATTTATCATC 1176
|||
Db 374 ValTyrGluGluGlnGlyThr-----SerLysValPheGlnGlyAsnLeuAspAsn 391
QY 1117 CAGGATGTCGTAATAACTTTTGGCCACCAATTAATTCACGTTTATTAGAGTGAATCCT 1236
|||
Db 392 SerHisLysLysAsnIlePheGluLysProPheMetAlaArgTyrValArgValLeuPro 411
QY 1237 ACCCAATGGCAGCAAAATTTGCCATGAATGGAGCTGCTCGGATGT 1284
|||
Db 412 LeuSerTrpHisAsnArgIleThrLeuArgLeuGluLeuLeuGlyCys 427
RESULT 10
A36479
milk fat globule membrane protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A36479
R:Stubbs, J.D.; Lekutis, C.; Singer, K.L.; Bul, A.; Yuzuki, D.; Srinivasan, U.; Parry
Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
A:Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the
A:Reference number: A36479; MUID:91046008; PMID:2122462
A:Accession: A36479
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-463 <STU>
A:Cross-references: GB:M38337; MID:gl9142; PIDN:AAA39534.1; PID:gl9143
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F:28-60/Domain: EGF homology <EGL>
F:68-107/Domain: EGF homology <EG2>
F:147-303/Domain: discoidin I amino-terminal homology <DN1>
F:307-463/Domain: discoidin I amino-terminal homology <DN2>
Alignment Scores:
Pred. No.: 4,71e-16 Length: 463
Score: 323.50 Matches: 114
Percent Similarity: 37.98% Conservative: 55
Best Local Similarity: 25.62% Mismatches: 143
Query Match: 7.81% Indels: 133
DB: 1 Gaps: 16
US-10-060-830-1 (1-2280) x A36479 (1-463)
QY 29 CCCTCCCGCTCCCTCCCTCCCTGCTGCTCAACTCCTCCTCTCCTCTCCATGCCTCTCTTCC 88
|||
Db 127 ProAsnThrAlaValProThrProAlaProThrProAspLeuSerAsnLeuAlaSer 146
QY 89 TCCTGCTCTACTGTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 148
|||
Db 147 ArgCysSer-----ThrGlnLeuGlyMetGluGlyGlyAlaIleAla 160
QY 149 GAT-----GTGGACACACTGTACTAGGCCCTGAGAGTGGACCCCTTACATCATAA 202
|||
Db 161 AspSerGlnIleSerAlaSerTyr----- 168
QY 203 ACCCAGACACTATCCCAACAGCAGCTGTTTGAATGGGAGATCCCTGTAAAGATGGGAG 262
|||
Db 169 -----ValTyrMetGlyPheMetGlyLeuGlnArgTrpGly 180
QY 263 AGAGAGTTCGCATCAAAATTTGGTGACTTTGACATTCAGATTCGATTCTTGTCACTTTA 372
|||
Db 181 ProGluLeuAla-ArgLeuTyr-----ArgThrGlyIleValAsnAla--- 194
QY 323 ATTACTTGAGAAATTTAATGAAT-----TGGAGTCAGCA 358
```


QY 573 AGCTGGTGTCTGCTCTCTCTGCTGATATCTGGAACAATT----- 615
Db 1849 olusLeuAlaArgLeuHisTyr-----SerGlySerIleAsnAlaTrpSerThrLy 1866
QY 616 ---CCTCATGATATAGAGATTCTCCCAATTGTGTGATGCTGCTGTCATGTCAGAGAGT 671
Db 1866 saspprohissertrpilleLysValAspLeuAlaPrometIleHisGlyIleMe 1886
QY 672 AGTGTCACAAACAGTGGCGCGCAATCAGTGTGTATTAGTAAAGGTATTCCTCTATTA 731
Db 1886 tthrGlnGlyAlaArgGlnLysPheSerSerLeuTyrIleSerGlnPheIleIleMetTy 1906
QY 732 TGAA----- 735
Db 1906 tserLeuAspGlyArgAsnTrpGlnSerTyrArgGlyAsnSerThrGlyThrLeuMetVa 1926
QY 736 -----AGTTCCTTGCTAACACGTC----- 756
Db 1926 lPhePheGlyAsnValAspAlaSerGlyIleLysHisAsnIlePheAsnProProIleVa 1946
QY 757 -----ACATCTGTGTGGGACACTTATCTACAAG 785
Db 1946 lAlaArgTyrIleArgLeuHisProThrHisTyrSerIleArgSerThrLeuArgMetGl 1966
QY 786 TCTTTTACATTAAGACAAGTGGATGTATGGAACACTGGGATGGAGTCTGGTGTGAT 845
Db 1966 uLeuMetGlyCysAspLeuAsnSerCysSerMetProLeuGlyMetGlnAsnLysAlaIl 1986
QY 846 CCGGATCCTCAATAACAGCATCATCTGTGTGAGTGGAGTGGACTGACACACAGGCGACA 905
Db 1986 eSerAspSerGlnIleThrAlaSerHisLeuSerAsnIlePheAlaThr----- 2003
QY 906 GAACAGTTGGAACCCAAACAGCGCTGAAAAAACCTGGA-----CGCCCTTGGGC 959
Db 2004 -----TrpSerProSerGlnAlaArgLeuHisLeuGlnGlyArgThrAsnAlaTrpAr 2021
QY 960 TCGTTTGGCACTGATGAATACCACTGGTTACAAATAGATTGTAAGAGAAAGAAAT 1019
Db 2021 gProArgValSerAlaGluGluTrpLeuGlnValAspLeuGlnLysThrValLysVa 2041
QY 1020 ACAGGCATTATAACCACTGGATCCACCATGGTGAGACAAATTAATGCTGTCCTA 1079
Db 2041 lThrGlyIleThrThrGlnGlyValLysSerLeuLeuSerSerMetTyrValLysGluPh 2061
QY 1080 CAGATCCCTGACAGTGATGGGAGAAATGGACTGTGTACAGAGACCTGGTGGA 1139
Db 2061 eLeuValSerSerGlnAspGlyArgArgTrpThrLeuPheLeuGlnAspGly----- 2079
QY 1140 GCAAGATAAGATATTTCAAGAAACAAAGATTATCACCAGGATGTGCGTAATAACTTTT 1199
Db 2080 -HisThrLysValPheGlnGlyAsnAspSerSerThrProValValAsnAlaLeuAs 2099
QY 1200 GCCACCAATTATGACGTTTATTATAGCTGAATCCTACCAATGGCAGCAGAAATGTC 1259
Db 2099 pProProLeuPheThrArgTyrLeuArgIleHisProThrSerTrpAlaGlnHisIleAl 2119
QY 1260 CATGAAATGGAGCTCTCGATGTCAG 1287
Db 2119 aLeuArgLeuGluValLeuGlyCysGlu 2128

RESULT 12
A47004
coagulation factor VIII precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 18-Jun-1999
C:Accession: A47004
R:Elder, B.; Lakich, D.; Gitschier, J.
Genomics 16, 374-379, 1993
A:Title: Sequence of the murine factor VIII CDNA.
A:Reference number: A47004; MUID:93300511; PMID:8314577
A:Accession: A47004
A>Status: preliminary

A:Molecule type: mRNA
A:Residues: 1-2319 <ELD>
A:Cross-references: GB:105573; NID:g192456; PIDN:AAA37385.1; PID:g192457
C:Superfamily: coagulation factor VIII; discoildin I amino-terminal homology; ferroxiid
F:1-19/Domain: signal sequence #status predicted <SIG>
F:23-349/Domain: ferroxiidase repeat homology <F01>
F:402-730/Domain: ferroxiidase repeat homology <F02>
F:1686-2006/Domain: ferroxiidase repeat homology <F03>
F:2007-2156/Domain: discoildin I amino-terminal homology <DN1>
F:2160-2313/Domain: discoildin I amino-terminal homology <DN2>

Alignment Scores: 6.24e-15 Length: 2319
Pred. No.: 310.00 Matches: 96
Score: 44.76% Conservative: 45
Percent Similarity: 30.48% Mismatches: 113
Best Local Similarity: 7.48% Indels: 61
Query Match: 2 Gaps: 11
DB: 11

US-10-060-830-1 (1-2280) x A47004 (1-2319)

QY 381 TGGTCTGGGTGCAAAATGAACCAATTCATTAAGCAAGCAATGAATCACAATTCT 440
Db 2049 TrpSer-----ThrLysGluProPheSerTrpIle-LysValAspLeuAlaProMe 2066
QY 441 GTTCATGAGTGGATCCATGTTCTGGA-----CG 470
Db 2066 tIleValHisGlyIleLysThrGlnGlyAlaArgGlnLysPheSerSerLeuTyrIleSe 2086
QY 471 CGGATTTTGGCCCTCATACTCTGTATAGATAAACAAGATCTAATTAATTTTGGACAC 530
Db 2086 rGlnPheIleIleMetTyrSerLeuAspGlyLysLys-----TrpLeuSerTy 2102
QY 531 TGCATCAAAATTTTGGAAACCTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTCTTCC 590
Db 2102 rGlnGlyAsn-----SerThrGlyThrLeuMetVa 2112
QY 591 TTTTGTCTGAGATATCTGGAACAATTCCTCATGGATATAGAGATTCTCTGCCAATGTGCAT 650
Db 2112 lPhe-----PheGlyAsnValAspSerSer----- 2120
QY 651 GGCTGGTGTGATGTCAGGATAGTCTCAACACAGTTGGGCGCCAAACAGTGTGTGTAAT 710
Db 2121 -----GlyIleLysHisAsnSerPheAsnProProIle-----Il 2132
QY 711 TAGTAAAGGTATTCCTTATTATGAAGTCTTTGGCTAACACAGTCACATCTGTGTGGG 770
Db 2132 eAlaArgTyrIleArgLeuHisProThrHis-----SerSerIleArgSe 2147
QY 771 ACACATATCTACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACTGGGGAT 830
Db 2147 rThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerIleProLeuGlyMe 2167
QY 831 GGAGTCTGGTGTGTCGGGATCTCAATAACACAGATCATCTGTCTGGAGTGGACTGA 890
Db 2167 tGluSerLysValIleSerAspThrGlnIleThrAlaSerSerTyPhe----- 2183
QY 891 CCACACAGGCGAAGAGACAGTGGAAACCCAAACAGCCAGCTGAAAAACCTGGA-- 948
Db 2184 -----ThrAsnMetPheAlaThrTrpSerProSerGlnAlaArgLeuHisLeuGlnGlyAr 2202
QY 949 -----CCGCTTGGGCTGCTTTTCCCACTGATGAATACCACTGTTTACAATAGATTGAA 1004
Db 2202 gThrAsnAlaTrpArgProGlnValAsnAspProLysGlnTrpLeuGlnValAspLeuGl 2222
QY 1005 TAAGGAAAGAAAATAACAGGCAATTATACCACTGGATCCACCTGGTGGAGCACAATTA 1064
Db 2222 nLysThrMetLysValThrGlyIleIleThrGlnGlyValLysSerLeuPheThrSerMe 2242
QY 1065 CTATCTGTCTGCTACAGAAATCCTGTACAGTGTATGATGGCAAGAAATGCTGTGTACAG 1124
Db 2242 tPheValLysGluPheLeuIleSerSerGlnAspGlyHisHisTrpThrGlnIleLeu 2262

Qy	1125	AGAGCCTGTGGTGGACCAAGATAAAGATATTTTCACAGGAACAAGATATATCACCAGAGTGT	1183
Db	2262	uTyraNGly-----LysValIysValPheGlnGlnAsnGlnAspSerSerThrProPhe	2280
Qy	1185	GCGTAAACTTTTGGCCACCAATATTGCGAGTTTATTAGACTGGATCCTACCCAATG	1244
Db	2280	tMetAsnSerLeuAspProLeuLeuThrArgTyrLeuArgIleHisProGlnIleTrp	2300
Qy	1245	GCACGAGAAAATGGCATGAAAATGGAGCTGCTGGATGTGCAG	1287
Db	2300	pGluHisGlnIleAlaLeuArgLeuGluIleLeuGlyCysGlu	2314
RESULT 13			
A44258			
factor VIII-associated gene B hypothetical protein - human			
C/Species: Homo sapiens (man)			
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999			
C/Accession: A44258			
R/Levinson, B.; Kenwick, S.; Gamel, P.; Fisher, K.; Gitschler, J.			
Genomics 14, 585-589, 1992			
A/Title: Evidence for a third transcript from the human factor VIII gene.			
A/Reference number: A44258; MUID:93052386; PMID:1427887			
A/Accession: A44258			
A>Status: preliminary			
A/Molecule type: mRNA			
A/Residues: 1-216 <LEV>			
A/Cross-references: GB:M90707; NID:g192316; PIDN:AAA58466.1; PID:g182317			
C/Superfamily: coagulation factor VIII; discoidin I amino-terminal homology			
F;1-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>			
F;57-210/Domain: discoidin I amino-terminal homology <DN2>			
Alignment Scores:			
Pred. No.:		5,47e-15	Length: 216
Score:		308.50	Matches: 66
Percent Similarity:		54.49%	Conservative: 31
Best Local Similarity:		37.08%	Mismatches: 72
Query Match:		7.45%	Indels: 9
DB:		2	Gaps: 3
US-10-060-830-1 (1-2280) x A44258 (1-216)			
Qy	760	TCGTGTGGTGGACACATTACTACAAGTCTTTTACATTTAAGACAAGTGGATGTATGGA	819
Db	41	SerIleArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet	60
Qy	820	ACACTGGGATGAGTCTGGTGTGATCGCGGATCTCCAATAACAGCATCATCTGCTGCTG	879
Db	61	ProLeuGlyMetGluSerLysLysAlaIleSerAspAlaGlnIleThrAlaSerTyrPhe	80
Qy	880	GAGTGGACTGACCACACAGCGGCAAGAGAACAGTCTGGAAACCCAAAAGCCAGGCTGAAA	939
Db	81	-----ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis	95
Qy	940	AAACCTTGA-----CCGCCTTTGGCTGCTTTTGGCATGTGAATACCACTGGTTACAA	993
Db	96	LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGln	115
Qy	994	ATAGATTGGAATGAAGAAAAGAAATACACGCATTATACCACTGGATCCACCATGGT	1053
Db	116	ValAspPheGlnLysThrMetLysValThrGlyValThrThrGlnGlyValLysSerLeu	135
Qy	1054	GAGCACAATTACTATATGTCTGCCACAGAATCCCTGTACAGTGTATGGTGGCAGAAATGG	1113
Db	136	LeuThrSerMetTyrValLysGluPheLeuIleSerSerSerGlnAspGlyHisGlnTrp	155
Qy	1114	ACTGTGTACAGAGAGCCTGGTGTGGACGACGATAATATTTCAAGGAAACAAAGATAT	1173
Db	156	ThrLeuPhePheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer	173
Qy	1174	CACGAGATGTGCGTAAATACATTTTTTGGCCACCAATATTGCACGCTTTTATTAGACTGAAT	1233
Db	174	PheThrProValValAsnSerLeuAspProLeuLeuThrArgTyrLeuArgIleHis	193

Qy 1234 CCTACCCAAATGGCAGCAGAAAATTCGATGCAATGAGCTGTCGGATGTCAG 1287
||| ||| :::::::::::::::::::::
Db 194 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 211

RESULT 14

EZHJ

coagulation factor VIII precursor [validated] - human
N:Alternate names: antihemophilic factor A; coagulant factor VIIc; procoagulant co
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence,revision 28-Aug-1985 #text_change 08-Dec-2000
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445;
R:Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A:Title: Sequence of the exon-containing regions of the human factor VIII gene.
A:Reference number: I54318; MUID:93265012; PMID:1303178
A:Accession: I54318
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1921, 'S', 1923-2351 <RES>
A:Cross-references: GB:M88648; NID:g182381; PIDN:AAA52420.1; PID:g182383
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Seeb
Nature 312, 330-337, 1984
A:Title: Expression of active human factor VIII from recombinant DNA clones.
A:Reference number: A00525; MUID:85061548; PMID:6438526
A:Accession: A00525
A:Molecule type: mRNA
A:Residues: 1-2351 <WOO>
A:Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179
R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Soltzman, L.A.; Buecker, J.L.; Pittman, D.
S. D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A:Reference number: I58059; MUID:85061550; PMID:6438528
A:Accession: I58059
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-74, 'V', 76-1259, 'E', 1261-2351 <RE2>
A:Cross-references: GB:X01740; NID:g182802; PIDN:AAA52484.1; PID:g182803
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; J.
B.; Randall, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.; J.
DNA 4, 333-349, 1985
A:Title: Characterization of the polypeptide composition of human factor VIII:C and t
A:Reference number: A23584; MUID:86081164; PMID:39355400
A:Accession: A23584
A:Molecule type: mRNA
A:Residues: 1-2351 <TRU>
A:Cross-references: GB:M14113; NID:g182817; PIDN:AAA52485.1; PID:g182818
R:Eaton, D.; Rodriguez, H.; Vehar, G.A.
Biochemistry 25, 505-512, 1986
A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleavag
ity.
A:Reference number: A26174; MUID:86159740; PMID:3082357
A:Accession: A26174
A:Molecule type: protein
A:Residues: 20-367;392-399, 'X', 401-402;1668-1678;1709-1722, 'D', 1723-1725;1741-1755 <EA
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.
Biochemistry 31, 3315-3325, 1992
A:Title: Identification and functional importance of tyrosine sulfate residues within
A:Reference number: A42348; MUID:92207952; PMID:1554716
A:Accession: A42348
A:Molecule type: protein
A:Residues: 20-367;356-371;392-408;582-594;1668-1669, 'X', 1671;1672-1692;1693-1708;1709
A:Experimental source: recombinant material from Chinese hamster ovary cells
A:Note: sequence extracted from NCBI backbone and corrected to correspond with the pu
R:Pay, P.J.; Smudzin, T.M.
J. Biol. Chem. 264, 14005-14010, 1989
A:Title: Intersubunit fluorescence energy transfer in human factor VIII.
A:Reference number: A43986; MUID:89340500; PMID:2503509
A:Accession: A43986
A:Molecule type: protein
A:Residues: 'X', 517-523;1853-1860, 'X', 1862-1864, 'X', 1866 <FAY>
R:Lytle, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens,
J. Biol. Chem. 266, 740-746, 1991

A:Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for
A:Reference number: A56109; MUID:91093266; PMID:1898735
A:Contents: annotation; sulfation
R:Gitschier, J.; Wood, W.I.; Goralika, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar,
Nature 312, 326-330, 1984
A:Title: Characterization of the human factor VIII gene.
A:Reference number: A56196; MUID:85061547; PMID:6438525
A:Contents: annotation; introns
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
Protein Sci. 4, 740-746, 1995
A:Title: Locations of disulfide bonds and free cysteines in the heavy and light chains of
A:Reference number: A56216; MUID:95338127; PMID:7613471
A:Contents: annotation; disulfide bonds
A:Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls
R:Kjalker, M.; Heding, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
Eur. J. Biochem. 234, 773-779, 1995
A:Title: Amino acid residues 721-729 are required for full factor VIII activity.
A:Reference number: S63527; MUID:96163459; PMID:8575434
A:Accession: S63527
A:Molecule type: protein
A:Residues: 733-752/753-759 <KJA>
R:Lind, P.; Larsson, K.; Spira, J.; Sydow-Baackman, M.; Almstedt, A.; Gray, E.; Sandberg
Eur. J. Biochem. 232, 19-27, 1995
A:Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
A:Reference number: S66445; MUID:96048024; PMID:7556150
A:Accession: S66445
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1668-1685 <LJN>
C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure to
C:Genetics:
A:Gene: GDB:F8C
A:Cross-references: GDB:119124; OMIM:306700
A:Map position: Xq28-Xq28
A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro
A:Pathway: blood coagulation
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-2351/Product: coagulation factor VIII #status experimental <NAV>
F:20-740/Product: coagulation factor VIIIA heavy chain #status experimental <ACH>
F:20-356/Domain: A1 <DA1>
F:23-348/Domain: ferroxidase repeat homology <FOL>
F:392-759/Domain: A2 <DA2>
F:402-730/Domain: ferroxidase repeat homology <FOL>
F:760-1667/Domain: B <DBO>
F:1668-2351/Product: coagulation factor VIIIA light chain #status experimental <ACL>
F:1709-2038/Domain: A3 <DA3>
F:1716-2038/Domain: ferroxidase repeat homology <FOL>
F:2039-2191/Domain: C1 <DC1>
F:2039-2188/Domain: discoidin I amino-terminal homology <DN1>
F:2192-2351/Domain: C2 <DC2>
F:2192-2345/Domain: discoidin I amino-terminal homology <DN2>
F:60-258,601,776,803,847,919,962,1020,1024,1074,1085,1204,1274,1278,1301,1319,1403,1
F:172-198,267-348,547-573,649-730,1851-1877,1918-1922,2040-2188/disulfide bonds: #status
F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted
F:365-737,738,742,1683,1699/Binding site: sulfate (Tyr) (covalent) #status experimental
F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:414,426/Binding site: sulfate (Tyr) (covalent) #status predicted
F:759-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental
F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental
F:2193-2345/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.: 8,09e-15 Length: 2351
Score: 308.50 Matches: 66
Percent Similarity: 54.49% Conservative: 31
Best Local Similarity: 37.08% Mismatches: 72
Query Match: 7.45% Indels: 9

DB: 1 Gaps: 3
US-10-060-830-1 (1-2280) x EZHU (1-2351)
Qy 760 TCTGTGTGGACACTTATCTACAGACTCTTTTACATTTAAGACAAAGTGGATGTTATGGA 819
Db 2176 SerileArgSerThrLeuArgMetGluMetGlyCysAspLeuAsnSerCysSerMet 2195
Qy 820 ACATGGGGATGGAGTGTGGTGTGATCGGGATCTCTCAATTAACACCATCATCTCTGTG 879
Db 2196 ProLeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyrPhe 2215
Qy 880 GAGTGGACTGACACAGAGCGGACAGAGAACAGTTCGGAACCCAAACCCAGGCTGAAA 939
Db 2216 -----ThrasmetPheAlaThrTrpSerProSerLysAlaArgLeuHis 2230
Qy 940 AAACCTGGA-----CCOCCTTGGGCTGCTTTTGGCCACTGATGAATACACAGGTTACAA 993
Db 2231 LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnProLysGluTrpLeuGln 2250
Qy 994 ATAGATTGATAGGAAAGAAATAACAGCAATTAACACCTGGATCCACCATGGTG 1053
Db 2251 ValAspPheGlnLysThrMetLysValThrGlyValThrThrGlnGlyValLysSerLeu 2270
Qy 1054 GAGCACATTAATGCTGCTCCCTACAGAACTCTACAGATGATGGGACAGAAATGG 1113
Db 2271 LeuThrSerMetTyrValLysGluPheLeuIleSerSerSerGlnAspLysGlnTrp 2290
Qy 1114 ACTGTGTACAGAGCCCTGGTGTGGAGCAAGATAAGATATTTCAAGGAACAAAGATTAT 1173
Db 2291 ThrLeuPheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer 2308
Qy 1174 CACAGAGATGGCGTAATAAAGTCTTTTGGCCACCAATTTATTCAGCTTTTATAGATGAAT 1233
Db 2309 PheThrProValValAsnSerLeuAspProProLeuLeuThrArgTyrLeuArgIleHis 2328
Qy 1234 CCTACCAATGCGACAGAAATTCCTCAAGAAATGGAGCTCTCGGATGTCAG 1287
Db 2329 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 2346
RESULT 15
T08618
intrinsic factor-B12 receptor CUBILIN precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Aug-2002
C:Accession: T08618
R:Moestrup, S.K.; Kozyraki, R.; Kristiansen, M.; Kaysen, J.H.; Rasmussen, H.H.; Braul
J. Biol. Chem. 273, 5235-5242, 1998
A:Title: The intrinsic factor-vitamin B12 receptor and target of teratogenic antibodi
A:Reference number: 216459; MUID:98148073; PMID:9478979
A:Accession: T08618
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3623 <MOE>
A:Cross-references: EMBL:AF022247; MUID:93834379; PIDN:AAC71661.1; PID:93834380
C:Genetics:
A:Gene: CUBILIN
C:Superfamily: intrinsic factor-B12 receptor cubilin; EGF homology
C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membra
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-3623/Product: intrinsic factor-B12 receptor CUBILIN #status predicted <MAT>
F:133-164/Domain: EGF homology <EGF>
F:436-467/Domain: EGF homology <EGF>
Alignment Scores:
Pred. No.: 1.14e-08 Length: 3623
Score: 226.50 Matches: 60
Percent Similarity: 41.63% Conservative: 37
Best Local Similarity: 25.75% Mismatches: 84
Query Match: 5.47% Indels: 52
DB: 2 Gaps: 7
US-10-060-830-1 (1-2280) x T08618 (1-3623)

Result No.	Query			ID	Description
	Score	Match	Length		
1	421.5	10.2	925	1	NRP2_RAT
2	412.5	10.0	931	1	NRP2_MOUSE
3	401.5	9.7	923	1	NRP1_HUMAN
4	398.5	9.6	931	1	NRP2_HUMAN
5	392.5	9.5	923	1	NRP1_MOUSE
6	388.5	9.4	922	1	NRP1_RAT
7	384.5	9.3	914	1	NRP1_CHICK
8	383	9.2	928	1	NRP1_XENLA
9	362	8.7	2258	1	FA5_PIG
10	361	8.7	2211	1	FA5_BOVIN
11	349.5	8.4	427	1	MFGM_BOVIN
12	347	8.4	2224	1	FA5_HUMAN
13	343	8.3	387	1	MFGM_HUMAN
14	340	8.2	409	1	MFGM_PIG
15	339	8.2	427	1	MFGM_RAT
16	323.5	7.8	463	1	MFGM_MOUSE
17	311.5	7.5	2133	1	FA8_PIG
18	310	7.5	2319	1	FA8_MOUSE

[illegible]

```
Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerThrArgGluTyrLeuTrp 460
QY 1405 -----ACGCACCACTACAACTCGCAGTACGATGAATTCCTCCGACAGACAGAA 1455
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTyrPheProArgAsnProGln 480
QY 1456 CAACAACCTGCAGTCTGTATCAGAAATACCTACCGTAATCCAAATGTAACCAAGAT 1515
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----CGCGTGGTGGCAGTCTTGTCTCCCTGTGCTGG 1548
Db 501 ValIleLeuGlnGlyAlaArgGlyAspSerIleThrAlaMetGluAlaArgAlaPhe 520
QY 1549 GTCATGGTCCCTACTCTCATCTCATATAGTGTGCTGTGGCACTGG----- 1599
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1600 -----AGAAACAGAAAGAAAGAAAGTGAAGGCACC-----TATGACTTACTCTTACTGG 1647
Db 541 ProArgThrGlnGlnProLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1648 GACCGGCGAGTGTGTGAAAGGAATGAAGCAGTTTCTTCTCGCAAAAGCACTGGACCAT 1707
Db 559 -----AspIleArgArgPhe----- 563
QY 1708 GAGGAAACCCAGTTCGCTATAGCAGCAGCGAAGTAATCACTGACTGATCCACAGAGAATC 1767
Db 564 GluProValProAlaGlnTyrValArgValTyrProGluArgTyrSerProAlaGlyIle 583
QY 1768 ACCACAGTGCCTCAGCTGCTCTCCAGAGTAT-----GCTCAGCCACTGTAGGAGGA 1821
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1822 ATTGTGTGACTACTCATCAAGATCTACCTTAAACCA-----GAAGAAGGA 1869
Db 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAAGAAGCAGGCTAT-----CGACCTAGATCTTCAACTCA 1908
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspAspLysAspLeuGln-LeuProSerG1 643
QY 1909 CAGGGCAGGAGTATATCATCCATGCTGAACCACTCCCAATACGGGCGCTAGTAT 1968
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTyrMetCysArgAlaLys 663
QY 1969 GCAACCCCAATCATCATGACATGTGACGGGCACCCCA----- 2005
Db 663 strLeuGlnSerThrTrpIleSerSerAlaAsnProAsnAspArgThrPheProAspAs 683
QY 2006 -----CAACTTCAG-----TTGGT 2019
Db 683 pLysAsnPheLeuLysLeuGlnSerAspGlyArgGluGlyGlnPheGlyArgLeu1 703
QY 2020 CAGCCCTCCACATCCACTTTCAGGCTACGGGACCAACCTCCCCCACTAGTGGGA 2079
Db 703 eSerProPro-ValHisLeu-----ProArgSerProValCysMetGluP 718
QY 2080 TACAATACACTTCTCTC-----C 2097
Db 718 heGlnTyrGlnAlaMetGlyHisGlyValAlaLeuGlnValValArgGluAlaArg 738
QY 2098 AGAGTACAGAGTGTCTC-----CTCAGCCCGAGCCCATGATGATACCCGGAAGCTGGG 2151
Db 738 InGluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlySerGluTyrLysHisGly 757
QY 2152 AACCCAGGTCTACCTCCCGACAGCAATGTTGCTACCAAGTGCACAGCAGCACACAGAA 2211
Db 758 ArgIleLeuProSerTyrAsp-----MetGluTyrGlnIle-----ValPheGluGly 774
QY 2212 GTATCAGGACGAGGAAGGATGGGAATGTGATGT 2247
Db 775 ValIleGlyLysGlyArgSerGlyGluIleSerIle 786
```

```
RESULT 2
NRP2_MOUSE
ID NRP2_MOUSE STANDARD: PRT: 931 AA
AC O35375; O35373; O35374; O35376; O35377; O35378;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165
DE receptor 2).
GN NRP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=BA1B/C;
RX MEDLINE=97470888; PubMed=9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III.";
RL Neuron 19:547-559(1997).
CC -!- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC -!- SUBUNIT: NEUROPILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC NEUROPILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN E RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -!- ALTERNATIVE PRODUCTS: 6 ISOFORMS; A22 (SHOWN HERE), A0, A5, A17,
CC B0 AND B5; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING CNS, PNS AND IN SOME
CC NONNEURAL TISSUES INCLUDING LIMB BUDS, DEVELOPING BONES, MUSCLES,
CC INTESTINAL EPITHELIUM, KIDNEY, LUNG AND SUBMANDIBULAR GLAND.
CC -!- DEVELOPMENTAL STAGE: THE EXPRESSION PATTERN IS VERY DYNAMIC AND
CC IS DEVELOPMENTALLY REGULATED.
CC -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF022856; AAC53379.1; -
CC EMBL; AF022854; AAC53377.1; -
CC EMBL; AF022855; AAC53378.1; -
CC EMBL; AF022857; AAC53380.1; -
CC EMBL; AF022858; AAC53381.1; -
CC EMBL; AF022861; AAC53382.1; -
CC HSSP; P12259; 1C2T.
CC MGD; MGI:1100492; Nrp2.
CC InterPro; IPR000859; CUB_domain.
CC InterPro; IPR000421; FA58_C.
CC InterPro; IPR000998; MAM_domain.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00629; MAM; 1.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00042; CUB; 2.
CC SMART; SM00231; FA58C; 2.
CC SMART; SM00137; MAM; 1.
CC PROSITE; PS01180; CUB; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PROSITE; PS00060; MAM_2; 1.
CC Transmembrane; Glycoprotein; Neuropeptide; Receptor;
KW Alternative splicing.
FT SIGNAL 1 20
FT POTENTIAL.
```


Db 441 LeuSerGlyLeuIleAlaAspThrGlnIleSerAlaSerSerThrArgGluTyrLeuTrp 460
QY 1405 -----ACGCAACCACTACACCTCGCAGTAGCAATTCCTCGACAGACAGAA 1455
Db 461 SerProSerAlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgAsnProGln 480
QY 1456 CAACAACACTGCCAGTCTGATATCAGAAATACCTACCGTAACATCAATGTAACCAAGAT 1515
Db 481 AlaGlnProGlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGly 500
QY 1516 GTA-----GCGCTGGCTCGAGTTCTTGTGCTCCCTGCTGCTG 1548
Db 501 ValIleIleGlnGlyAlaArgGlyGlyAspSerIleThrAlaValAlaArgAlaPhe 520
QY 1549 GTATGGTCTCTACTCTCTCATCTCATATTAGTGTGCTGGCACTGG----- 1599
Db 521 ValArgLysPheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAsp 540
QY 1600 -----AGAAACAGAAAGAAAGAAAGCAAGGCAACC-----TATGACTTACCTTACTTGG 1647
Db 541 ProArgThrGlnGlnThrLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1648 GACCGGCAGGTGGTGGAAAGGAATGAAGAGTTTCTCTGCAAAAGCAGTGGACCAT 1707
Db 559 -----AspIleArgArgPheAspPro----- 565
QY 1708 GAGGAACCCAGTTCCTATATAGCAGCAGCAAGTAAATCACTGAGTCCAGAGAGATC 1767
Db 584 GlyMetArgLeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThr 603
QY 1822 ATTCTTGGTACACTTCATCAAGATCATACCTTAAACA-----GAAGAAGCA 1869
Db 604 LeuGlyProThrValLysSerGluGluThrThrThrProTyrProMetAspGluAspAla 623
QY 1870 AAAGAAGCAGCTAT-----GCAGACCTAGATCTTCAACATCA 1908
Db 624 ThrGluCysGlyGluAsnCysSerPheGluAspLysAspLeuGln-LeuProSerG1 643
QY 1909 CCAGGCGAGGAAGTTTATCATGCCT-----ATGCTGAACCACTCCCA 1950
Db 643 yPheAsnCysAsnPheAspPheProGluGluThrCysGlyTrpValTyrAspHisAlaLys 663
QY 1951 A---TTAGGGGCGCTGAGTATGCAACCCCAATCATCATGTCAGTGGCAGCCCA-- 2005
Db 663 strPleuArg-----SerThrTrpIleSerSerAlaAsnProAs 676
QY 2006 -----CAACTTCAG----- 2014
Db 676 nAspArgThrPheProAspAspLysAsnPheLeuLysLeuGlnSerAspGlyArgArgG1 696
QY 2015 -----TTGGTCAGCCCTCCATCCATCCATCTTCAAGGCTACGGGGAACCAA 2058
Db 696 uGlyGlnTyrGlyArgLeuIleSerProPro-ValHisLeu-----ProA 711
QY 2059 CCTCCCCACTAGTGGGAGCAATCATACCTTCTCTC----- 2096
Db 711 rgSerProValCysMetGluPheGlnTyrGlnAlaMetGlyGlyHisGlyValAlaLeuG 731
QY 2097 -----CAGGACTCAGACTGCTC-----CTCAGCCAGCCGAC 2130
Db 731 InValValArgGluAlaSerGlnGlnSerLysLeuLeuTrpValIle-ArgGluAspGln 750
QY 2131 TATGATACCCGAAAGCTGGGAAGCCAGGTCTTACCTGCCAGGCAATGTGTACCA 2190
Db 751 GlySerGluTrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGln 769
QY 2191 GTGCCACAGACACAGAAGATATCAGAGCAGGAGGATGGGAA 2238
Db 770 Ile-----ValPheGluGlyValIleGlyLysArgSerGlyGlu 783

RESULT 3
NRPL_HUMAN
ID NRPL_HUMAN STANDARD; PRT; 923 AA.
AC OL4786; 060461.
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuropilin-1 precursor (Vascular endothelial cell growth factor 165 receptor).
GN NRPL OR NRP OR VEGF165R.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM).
RX MEDLINE=97433084; PubMed=9288753;
RA He Z., Tessier-Lavigne M.;
RT "Neuropilin is a receptor for the axonal chemorepellent semaphorin III."
RL Cell 90:739-751(1997).
RN [2]
RP SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM), AND SEQUENCE OF 22-39.
RC TISSUE=Breast;
RX MEDLINE=98188099; PubMed=9529250;
RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor."
RL Cell 92:735-745(1998).
RN [3]
RP SEQUENCE FROM N.A. (SOLUBLE/SNRPL ISOFORM), AND SEQUENCE OF 22-31.
RC TISSUE=Prostatic adenocarcinoma;
RX MEDLINE=20183929; PubMed=10688880;
RA Gagnon M.L., Bieleberg D.R., Gechtman Z., Miao H.-Q., Takashima S., Soker S., Klagsbrun M.;
RT "Identification of a natural soluble neuropilin-1 that binds vascular endothelial growth factor: In vivo expression and antitumor activity."
RL Proc. Natl. Acad. Sci. U.S.A. 97:2573-2578(2000).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid form of vascular endothelial growth factor (VEGF) and of placenta growth factor-2, but only neuropilin-2 functions as a receptor for the 145-amino acid form of VEGF."
RL J. Biol. Chem. 275:18040-18045(2000).
CC -1- FUNCTION: THE MEMBRANE-BOUND ISOFORM IS A RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGF, THE VEGF-165 ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS. IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS.
CC -1- FUNCTION: THE SOLUBLE/SNRPL ISOFORM BINDS VEGF-165 AND APPEARS TO INHIBIT ITS BINDING TO CELLS. IT MAY ALSO INDUCE APOPTOSIS BY SEQUESTERING VEGF-165. MAY BIND AS WELL VARIOUS MEMBERS OF THE SEMAPHORIN FAMILY. ITS EXPRESSION HAS AN AVERSE EFFECT ON BLOOD VESSEL NUMBER AND INTEGRITY.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. THE SOLUBLE/SNRPL ISOFORM IS SECRETED.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; MEMBRANE-BOUND (SHOWN HERE) AND SOLUBLE/SNRPL; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: THE EXPRESSION OF MEMBRANE-BOUND AND SOLUBLE/SNRPL ISOFORMS DOES NOT SEEM TO OVERLAP. THE MEMBRANE-BOUND ISOFORM IS EXPRESSED BY THE BLOOD VESSELS OF DIFFERENT TISSUES. IN THE DEVELOPING EMBRYO IT IS FOUND PREDOMINANTLY IN THE NERVOUS SYSTEM. IN ADULT TISSUES, IT IS HIGHLY EXPRESSED IN HEART AND PLACENTA; MODERATELY IN LUNG, LIVER, SKELETAL MUSCLE, KIDNEY

Db 245 SerSerLeuAlaAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr 264
 QY 721 TTTAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTATCGGGATCCT 780
 Db 265 PheLysThrSerGlyCysTyrglyThrLeuGlyMetGluSerGlyValIleAlaAspPro 284
 QY 781 CAATAACACATCATCTGTCTGGAGTGGACTGACACACAGGCGGCAAGAGACAGTTGG 840
 Db 285 GlnIleThrAlaSerSerValLeuGluThrPheAspHisThrGlyGlnGluAsnSerTrp 304
 QY 841 AAACCCAAAAAGCCAGGCTCAAAAAACCTGGACCGCTTGGCTGCTTTTGGCCACTGAT 900
 Db 305 LysProLysLysAlaGlyLeuLysLysProGlyProProTrpAlaAlaPheAlaThrAsp 324
 QY 901 GAATACCACTGGTTACAAATAGATTGAATAAGAAAGAAATACAGGCATTTAACC 960
 Db 325 GluTyrglnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleThr 344
 QY 961 ACTGGATCCCACTGGTGGAGCACAATTAATGTCTGCTGCTACAGAACTCCTGTACAGT 1020
 Db 345 ThrGlyIleThrMetValGluHisAsnTyrrValSerAlaTyrrArgIleLeuTyrrSer 364
 QY 1021 GATGATGGCAGAATCGACTGTGTACAGAGCGCTGGTGGAGCAAGATAAGATATTT 1080
 Db 365 AspAspGlyGlnLysTrpThrValTyrrArgGluProGlyValGluGlnAspLysIlePhe 384
 QY 1081 CAAGGAACAAGATTATCACCAGGATGTCGTAATAACATTTTGGCCACCAATATTGCA 1140
 Db 385 GlnGlyAsnLysAspTyrrHisGlnAspValArgAsn***PheLeuProIleIleAla 404
 QY 1141 CGTTTATTAGAGTAATCTACCAATGCGCAGCAGAGAAATGCCATGAAATGGAGCTG 1200
 Db 405 ArgPheIleArgValAsnProThrGlnTrpGlnLysIleAlaMetLysMetGluLeu 424
 QY 1201 CTCGGATGTCAGTTTATTCCTAAGAGTCGCTCCCTCCAAACCTTACTCAACCTCCACCTCCT 1260
 Db 425 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProPro 444
 QY 1261 CGGACAGCAATGACCTCAAAACACTACAGCCCTCCAAATAGGCAAGGTGCTGCC 1320
 Db 445 ArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla 464
 QY 1321 CCRAAATTAGCAACCACTACACCTCGCAGTAGCAATGAATTTCTCGCACACAGAA 1380
 Db 465 ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 484
 QY 1381 CAAACAACCTGCTGATATCAGAAATACCTACCTACCTCAAAATAGGCAAGGTGCTGCC 1440
 Db 485 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 504
 QY 1441 GTACGCTGGCTGAGTCTTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 Db 505 ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 524
 QY 1501 ATATTAGTGTGCTGGCTGGAGAAACAGAAAGAAAAAACTGAAGGCACCTATGAC 1560
 Db 525 IleLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrrAsp 544
 QY 1561 TTACTTACTGGGACGGGAGGTGGTGGGAAGGAATG 1599
 Db 545 LeuProTyrrTrpAspArgAlaGlyAsnSerArgGlyLeu 557

RESULT 3

AAE22716

ID AAE22716 standard; Protein: 385 AA.

XX AC

AAE22716;

XX AC

DT 09-AUG-2002 (first entry)

XX XX

Human neuropilin-Hy2 protein.

XX XX

Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation;
 neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
 thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
 ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
 wound healing; tissue repair; Parkinson's disease; Huntington's disease;
 amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
 cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
 human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
 systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
 autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
 myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
 neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic;
 cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic;
 immunosuppressive; chromosome 6q21.

Homo sapiens.

Key	Location/Qualifiers
Peptide	1..20
Protein	/label= signal_peptide
	61..385
	/note= "Human mature neuropilin-Hy2 protein"

WO200222815-A1.

21-MAR-2002.

12-SEP-2001; 2001WO-US28488.

11-SEP-2000; 2000US-0659671.

06-SEP-2001; 2001US-317902P.

(HYSE-) HYSEQ INC.

Tang YT;

WPI: 2002-393966/42.

N-PSDB; AAD35994.

Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
 useful for treating neurodegenerative diseases e.g. Alzheimer's
 disease, and for diagnosing and mapping genetic neuronal defects -

Claim 3; Page 128-130; 152pp; English.

The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
 neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
 like polypeptides and polynucleotides are useful in modulating neuronal
 growth regenerative capacity, treating neurodegenerative diseases,
 diagnosing and mapping genetic neuronal defects and degenerative diseases
 like Alzheimer's disease and for treating learning and memory disorders.
 They are also useful for inducing angiogenesis, neovascularisation, as
 well as organ growth and development e.g. heart and other tissues.
 Antagonists of neuropilin-like polypeptides are useful for treating
 cancers and other malignant diseases. Neuropilin is used to treat
 platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
 nocturnal haemoglobinuria and is used in nerve tissue growth or
 regeneration, in wound healing, tissue repair and replacement and in
 healing of bones, incisions and ulcers. Compositions comprising the
 sequences of the invention are useful for treating diseases of peripheral
 nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
 syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g.
 spinal cord disorders, head trauma and cerebrovascular diseases e.g.
 stroke, ulcers, immune deficiencies and immune disorders, infections by
 hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
 mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
 multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
 mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
 inflammatory eye disease. The nucleic acids of the invention are used in
 gene therapy techniques. The present sequence is human neuropilin-Hy2
 protein. Neuropilin-Hy2 gene is located on chromosome 6q21.

XX

for detecting full-matches or mismatches to the cDNAs. The genes for neuropilin-Hy1 and Hy2 are located on human chromosome 6q21. The nucleic acid array is useful for detecting full-matches or mismatches to the cDNAs. The neuropilin-like proteins and encoding cDNAs are useful in modulating neuronal growth, regenerative capacity, treating neurodegenerative diseases, learning and memory disorders, diagnosing and mapping genetic neuronal defects and degenerative diseases like Alzheimer's disease, for inducing angiogenesis, and neovascularisation and organ growth and development (e.g. the heart). The nervous system disorders include lesions of central or peripheral nervous systems, including traumatic lesions, ischaemic lesions, infectious lesions, degenerative lesions, lesions associated with nutritional diseases or disorders, neurological lesions, and lesions caused by toxic substances. The neuropilin-like proteins and cDNAs are also useful as markers for cancers. The neuropilin-like proteins are useful for regulating cell proliferation, cell differentiation, stem cell growth factor activity, for inducing proliferation of neural cells, regeneration of nerve and brain tissue, for treatment of central and peripheral nervous system diseases, and neuropathies, such as Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, to regulate haematopoiesis and treat myeloid and lymphoid cell disorders, various anaemias, and platelet disorders, such as thrombocytopenia, regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and as a food supplement or molecular weight marker. The cDNAs are useful in gene identification, genome mapping, transgenics, as hybridisation probes, for primer design, for gene chips and as a DNA antigen. The present sequence represents neuropilin-Hy2.

XX
SQ Sequence 385 AA;

Alignment Scores:

Pred. No.:	2.07e-160	Length:	385
Score:	1841.50	Matches:	372
Percent Similarity:	96.90%	Conservative:	3
Best Local Similarity:	96.12%	Mismatches:	8
Query Match:	46.40%	Indels:	7
DB:	23	Gaps:	1

US-10-060-830-2 (1-2190) x AAU79460 (1-385)

QY	1	ATGCCTCTGTTCCCTCTTACTTGCTGCTGCTCGAGGAGCGTGGAGCC	60
Db	1	MetProLeuPheLeuLeuLeuLeuValLeuLeuLeuLeuLeuGluAspAlaGlyAla	20
QY	61	CAGCAAGTGATGGATGGACACTGTACTAGGCCTGTAGAGTGGAAACCTTACATCC	120
Db	21	GlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSer	40
QY	121	ATAAACTACCACAGACCTATCCCAACAGCACTGTTTGTGAATGGGAGATCCGTGAAG	180
Db	41	IleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluIleArgValLys	60
QY	181	ATGGGAGAGAGATCCGATCAATTTGGTGACTTGACATTGAGATTCGTATCTGT	240
Db	61	MetGlyGluArgValArgIleLysPheGlyAspPheAspPheAspSerCys	80
QY	241	CACITTAATTAATCTGAGAAATTAATGGAATTTGGAGTGCAGCAGAACTGAATAGGCAA	300
Db	81	HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys	100
QY	301	TACTGTGCTGGGTTCCAAATGAACACCTCAATTAATCAAAAGGCAATGAATCACA	360
Db	101	TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr	120
QY	361	TTGCTGTTTCATGAGTGGAAATCATGTTTCTGGACGCGGATTTTGGCTCATCTCTGTT	420
Db	121	LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal	140
QY	421	ATAGATAACAGATCTAATTAATCTGTTTGGACACTGCATCCAAATTTTTCGAACCTGAG	480
Db	141	IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu	160

QY	481	TTCACTAAGTACTGCCAGCTGGTGTCTGCTCTCTTTTGTGTGAGATATCTGGAAACAATT	540
Db	161	PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIle	180
QY	541	CCTCATGGATATAGATTCCTCCCATTTGTCATGGTGGTGTGTCATCCAGGAGTACTG	600
Db	181	ProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValVal	200
QY	601	TCAAACACGTGGCGCGCAATCAGTCTTGTAAATAGTAAAGGTATTCCCTTATTATGAA	660
Db	201	SerAsnThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGlu	220
QY	661	AGTTCTTTGGCTAAACAGTCACATCTCTGTGGGACACTTATCTACAAAGTCTTTTACA	720
Db	221	SerSerLeuAlaAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr	240
QY	721	TTTAAGCAAGTGGATGTTATGGACACTGGGATGGAGTCTGGTGT -CATCGCGGATCC	779
Db	241	PheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyGlyAspArgGlySer	260
QY	780	TCAAATAACAGCATCATCTGTGCTGGAGTGGACTGACACACAGGCGCAAGAGACAGTTG	839
Db	261	SerAsnAsnSerIle-ThrValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTr	280
QY	840	GAACCCCAAAAGC -CAGGCTGAAAAACCTGGACCCCTGGGCTGCTTTTGGCACTG	898
Db	280	pLysProLysLysSerGlnAlaGluLysThrTrpThrAlaLeuGly-AlaPheAlaThrA	300
QY	899	ATGAATACCAGTGGTTACAAATAGATTGTAATAGAAAGAAATACAGGCATTATAA	958
Db	300	spGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysIleThrGlyIleIleIle	320
QY	959	CCACTGGATCCACCATGGTGGAGCACAATTACTATGTCTGCTCCCTACAGAATCTGTACA	1018
Db	320	hrThrGlySerThrMetVal-SerThrIleThrMetCysLeuProThrGluSerCysThr	339
QY	1019	GTGATGATGGCGAAATGGACTGTGTACAGAGACCTGGT -GTGGAGCAAGATGAATA	1077
Db	340	ValMetMetGlyArgAsnGlyLeuCysThrGluSerLeuValValGluGlnAspLysIle	359
QY	1078	TTTCAAGAAACAAGAT - -TATCACCAGGATGGCGTAACTATTTTGGCCACCAATT	1134
Db	360	PheGlnGlyAsnLysArgIleIleThrArgMetValArgAsnAsnPheLeuProIle	379
QY	1135	ATTGCACGTTTATT 1149	
Db	380	IleAlaArgLeuLeu 384	

RESULT 5
ABB97386
ID ABB97386 standard; Protein: 385 AA.
XX
AC ABB97386;
XX
XX 27-JUN-2002 (first entry)
XX
XX Novel human protein SEQ ID NO: 654.
XX
XX Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
XX
XX antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
XX
XX neuroprotective; antiparkinsonian; protein therapy; EST;
XX
XX expressed sequence tag.
XX
XX Homo sapiens.
XX
XX WO200222660-A2.
XX
XX 21-MAR-2002.
XX
XX 10-SEP-2001; 2001WO-US26015.
XX
XX 11-SEP-2000; 2000US-0659671.
XX

XX 12-SEP-2001; 2001WO-US28488.
XX 11-SEP-2000; 2000US-0659671.
PR 06-SEP-2001; 2001US-317902P.
XX (HYSE-) HYSEQ INC.
FA
XX
PI Tang YT;
XX WPI; 2002-393966/42.
XX Novel isolated human Neuropilin-Hyl and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects
XX
XX Disclosure; Page 131-132; 152pp; English.
XX
XX The invention relates to neuropilin-like polypeptide (neuropilin-Hyl and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.
CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat
CC platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal
CC nocturnal haemoglobinuria and is used in nerve tissue growth or
CC regeneration, in wound healing, tissue repair and replacement and in
CC healing of bones, incisions and ulcers. Compositions comprising the
CC sequences of the invention are useful for treating diseases of peripheral
CC nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager
CC spinal cord disorders, head trauma and cerebrovascular diseases e.g.
CC stroke, ulcers, immune deficiencies and immune disorders, infections by
CC hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,
CC mycobacteria, Leishmania spp., malaria spp., autoimmune disorders e.g.
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, autoimmune thyroiditis, diabetes
CC mellitus, graft-versus-host disease, myasthenia gravis and autoimmune
CC inflammatory eye disease. The nucleic acids of the invention are used in
CC gene therapy techniques. The present sequence is human neuropilin-Hy2
XX mature protein.
XX
SQ Sequence 365 AA;

Alignment Scores:
Pred. No.: 4,72e-152 Length: 365
Score: 1750, 50 Matches: 352
Percent Similarity: 96.73% Conservative: 3
Best Local Similarity: 95.91% Mismatches: 8
Query Match: 44.10% Indels: 7
DB: 23 Gaps: 1

US-10-060-830-2 (1-2190) x AAE22721 (1-365)

QY 61 CAGCAAGGTGATGGTGGACACACTGTACTAGCCCTGAGTGGAAACCCCTTACATCC 120
Db 1 GlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSer 20
QY 121 ATAAACTACCCACACACTATCCCAACAGCACTGTTCTGTAATGGGATCCGCTGAAG 180
Db 21 IleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluIleArgVallys 40
QY 181 ATGGGAGAGAGATTCGCATCAATTTGGTGACTTTGACATTTGAAGATTCGATCTTGT 240
Db 41 MetGlyGluArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCys 60
QY 241 CACTTTAATCTTGAGATTTATTAATGAATTTGGAGTCAGCAACACTGAATAGGCAAA 300
Db 61 HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLys 80

QY 301 TACTGTGCTCTGGGGTTGCAAAATGAACATTCAATTTGAATCAAAAGGCAATGAATCACA 360
Db 81 TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr 100
QY 361 TTGCTGTTCATGAGTGGAAATCCATGTTCTTCTGGACGGGATTTTGGCTCATCTGTT 420
Db 101 LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal 120
QY 421 ATAGATRAACAAGATCTAATTACTTGTGTGGACACTGCATCCAAATTTTTCGAACCTGAG 480
Db 121 IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu 140
QY 481 TTCAGTAAGTACTCCCGAGCTGGTGTCTCTCTCTTGTCTGAGATATCTGGAAACAAT 540
Db 141 PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIle 160
QY 541 CCTCATGATATAGAGATTCCTCCCATTTGTCATGCTGTCATGTCATGAGGAGTAGTG 600
Db 161 ProHisGlyTyrArgAspSerSerProLeuCysMetalaGlyValHisAlaGlyValVal 180
QY 601 TCAACACGTTGGCGGCCCAANTCAGTCTTCTAATTAGTAAGGTATTCCTATATGAA 660
Db 181 SerAsnThrLeuGlyGlyGlnIleSerValIleSerLysGlyIleProTyrTrpGlu 200
QY 661 AGTTCTTTGGCTAAACAACGTCACATCTCTGCTGGGACACTTATCTACAAAGTCTTTTACA 720
Db 201 SerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr 220
QY 721 TTTAAGCAAGTGGATGTTATGGAACACTGGGATGGAGTGTGTGTGT-GATCGGGATGCC 779
Db 221 PheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyGlyAspArgGlySer 240
QY 780 TCAATACACCATCATCTGCTGGAGTGGAGTGCACACACAGGCAAGGACAGACAGTTG 839
Db 241 SerAsnAsnSerIle-ThrValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTr 260
QY 840 GAAACCCCAAAAAGC-CAGGCTGAAAACCTGGACCCCTTGGGCTGCTTTTGGCACTG 898
Db 260 PLeysProLysLysSerGlnAlaGluLysThrTrpThrAlaLeuGly-AlaPheAlaThrA 280
QY 899 ATGAATACCACTGTTTCAAAATAGATTGTAATAGGAAAGAAATAACAGCATTATAA 958
Db 280 spGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleT 300
QY 959 CCACTGGATCCACCATGCTGGAGCACATTTACTATGCTGCTGCCTACAGATCCCTGTACA 1018
Db 300 hrThrGlySerThrMetVal-SerThrIleThrMetCysLeuProThrGluSerCysThr 319
QY 1019 GTGATGATGGCAAAATGGACTGTGTACAGAGAGCCCTGGT-GTGGCAAGATAAGATA 1077
Db 320 ValMetMetGlyArgAsnGlyLeuCysThrGluSerLeuValValGluGlnAspLysIle 339
QY 1078 TTTCAGGAACAAAGAT---TATCACAGGATGTCGCTAATACTAATCTTTTGGCCACAAT 1134
Db 340 PheGlnGlyAsnLysArgIleIleThrArgMetValArgAsnAsnPheLeuProIle 359
QY 1135 ATTGCACGTTTTAT 1149
Db 360 IleAlaArgLeuLeu 364
RESULT 7
AAU00670
ID AAU00670 standard; Protein; 715 AA.
XX
XX AC AAU00670;
XX
XX 07-SEP-2001 (first entry)
XX
XX Human TANGO 229 polypeptide.
XX
XX Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;

KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
 KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease;
 KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
 KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
 KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
 KW muscular dystrophy; immuno-competence; vertebrae; blood; serum.

XX Homo sapiens.

OS
 XX
 XX Location/Qualifiers
 FH Key
 FT Peptide
 FT 1..34
 FT /note= "signal peptide"
 FT Domain
 FT 35..455
 FT /note= "Extracellular domain"
 FT Protein
 FT 35..715
 FT /note= "Mature human TANGO 229"
 FT Domain
 FT 456..480
 FT /note= "Transmembrane domain"
 FT Domain
 FT 481..715
 FT /note= "Cytoplasmic domain"

XX WO200129088-A1.

XX 26-APR-2001.

XX 23-JUN-2000; 2000WO-US17386.

XX 19-OCT-1999; 99US-0420707.

XX (MILL-) MILLENNIUM PHARM INC.

XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

XX WPI; 2001-308477/32.

XX N-PSDB; AAS00660.

XX New isolated nucleic acid molecule for diagnosis, prevention, and
 PT therapy of human and other animal disorder, or as modulating agent for
 PT regulating cellular processes

XX Claim 8; Fig 1; 263pp; English.

XX The sequence represents human TANGO 229 polypeptide. This protein and
 CC similar others exhibit the ability to affect growth, proliferation,
 CC survival, differentiation, activity, morphology, or movement/migration
 CC of, e.g. T cells and cells of the heart, liver, pancreas, placenta,
 CC brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral
 CC blood leukocyte, bone marrow or thymus tissue. They can be used as
 CC modulating agents for regulating cellular processes, thus, the proteins
 CC and their associated nucleic acids can be used to prognosticate, prevent,
 CC diagnose, or treat disorders associated with physiological processes.
 CC These disorders include abnormal blood coagulation, asthma, anaphylaxis,
 CC hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria,
 CC atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention
 CC deficit disorder, Crohn's disease, gastroenteritis, goitre,
 CC hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and
 CC muscular dystrophy. Antibodies to disorders such as these can be made by
 CC providing a polypeptide of the invention to an immuno-competent
 CC vertebrate and harvesting blood or serum from the vertebrate.

XX Sequence 715 AA;

Alignment Scores:

Pred. No.:	8.75e-87	Length:	715
Score:	1042.00	Matches:	264
Percent Similarity:	52.14%	Conservative:	113
Best Local Similarity:	36.51%	Mismatches:	238
Query Match:	26.25%	Indels:	108
DB:	22	Gaps:	24

US-10-060-830-2 (1-2190) x AAU00670 (1-715)

QY 13 CTCCTGCTTACTTGCTCTCTGCTCGAGGACGCTGGAGCCGCAAGGTGAT 72

Db	20	LeuAlaLeuLeuLeuAlaValSerAlaProLeuArgLeuGlnAlaGluLeuGlyAsp	39
QY	73	GGATGTGCACACACTGTACTAGGCCCTGAGAGTGGACCCCTTACATCCATAAATACCCCA	132
Db	40	GlyCysGlyHisLeuValThrTyrGlnAspSerGlyThrMetThrSerLysAsnTyrPro	59
QY	133	CAGACCTATCCCAACAGCACTGTTTGTGAATGGAGATCCCTGTAAGATGGGAGAGAGA	192
Db	60	GlyThrTyrProAsnHisThrValCysGluLysThrIleThrValProLysGlyLysArg	79
QY	193	GTTCCGATCAAAATTTGGTGTGACATTTGACATTTGATTCCTTGTCTCACTTAAATAC	252
Db	80	LeuIleLeuArgLeuGlyAspLeuAspIleGlu---SerGlnThrCysAlaSerAspTyr	98
QY	253	TTGAGAATTTATATGGAATTTGGAGTCAGCAGAACTGAAATAGCAATACTCTGGCTG	312
Db	99	LeuLeuPheThrSer-----SerSerAspGlnTyrGlyProTyrCysGly---	113
QY	313	GGGTGCAAAATGAACCAATTCATTTGAATCAAAAGCAATGAATCAATTCCTGTTGATG	372
Db	114	SerMetThrValProLysGluLeuLeuLeuAsnThrSerGluValThrValArgPheGlu	133
QY	373	AGTGAATCCATGTTCTCTGGACGCGGATTTTGGCCTCATCTACTGTGTATAGATAACAA	432
Db	134	SerGlySerHisIleSerGlyArgGlyPheLeuLeuThrTyrAlaSerSerAspHisPro	153
QY	433	GATCTAATTTACTTGTGGACACTGCATCCAAATTTTGGAACTGAGTTCAGTAAGTAC	492
Db	154	AspLeuIleThrCysLeuGluArgAlaSerHisTyrLeuLysThrGluTyrSerLysPhe	173
QY	493	TGCCAGCTGGTGTCTGCTTCTCTTCTTCTGAGATATCTGCAACAATTCCTCATGGATAT	552
Db	174	CysProAlaGlyCysArgAspValAlaGlyAspIleSerGlyAsnMetValAspGlyTyr	193
QY	553	AGATATCTCTCCCATTTGTGCATGGCTGTGCATGCAGGAGTAGTGTCAACACCGTTG	612
Db	194	ArgAspThrSerLeuLeuCysLysAlaIleHisAlaGlyIleIleAlaAspGluLeu	213
QY	613	GGCGCCAAATCAGTGTGTAAATTAGTAAGGTATTCCTATTATGAAGTCTCTTGCT	672
Db	214	GlyGlyGlnIleSerValLeuGlnArgLysGlyIleSerArgTyrGluGlyIleAla	233
QY	673	AACACGTCACATCTGTGTGGACACTATCTACAAGTCTTTTACATTTAAAGCAAGT	732
Db	234	AsnGlyValLeuSerArgAspGlySerLeuSerAspLysArgPheLeuPheThrSerAsn	253
QY	733	GGATGTTATGGAACACACTGGGATGGAGTCTGTGTGTGTCGCGGATCTCTCAATAACAGCA	792
Db	254	GlyCysSerArgSerLeuSerPheGlu-----ProAspGlyGlnIleArgAla	269
QY	793	TCATCTGTGCTGGAGTGGACACTGACACACAGCGGCAAGAGACAGTGTGAACCCAAAAA	852
Db	270	SerSerSerTrpGlnSerValAsnGluSerGlyAspGlnValHisTrpSerProGlyGln	289
QY	853	GCCAGGCTGAAAAACCTGGACCGCTTGGCTGGCTGCTTTTGGCCACTGATGAA-----	903
Db	290	AlaArgLeuGlnAspGlnGlyProSerTrpAlaSerGlyAspSerSerAsnHisLys	309
QY	904	---TACCAGTGGTTACAAATAGATTTGAATAGGAAAAAGAAATAACAGGCATTATAACC	960
Db	310	ProArgGluTrpLeuGluIleAspLeuGlyGluLysLysIleThrGlyIleArgThr	329
QY	961	ACTGATCCACCATGGTGGACACAATTTACTATGTGTGTGCTGCTACAGAACTCTGTACAGT	1020
Db	330	ThrGlySerThrGlnSerAsnPheAsnPheTyrValLysSerPheValMetAsnHisLys	349
QY	1021	GATGATGGCAGAAATGGACTGTGTACAGAGAGCCCTGTGTGGAGCAAGATAAGATATT	1080
Db	350	AsnAsnAsnSerLysTrpLysThrTyrLysGlyIleValAsnAsnGluGluLysValPhe	369
QY	1081	CAAGGAAACAAAGATTATCACCAGGATGTGGTATATAACTTTTCCACCACTATTATTGCA	1140

Db 370 GlnGlyAsnSerAsnPheArgAspProValClnAsnAsnPhelieProProIleValala 389
Qy 1141 CGTTTTATTAGATGAATCTCCATCCCATGGCAGCAAAATTTGCCATGAATGGAGCTG 1200
|||:||||| ||| |||:|||||:|||||:|||||
Db 390 ArgTyrValArgValProGlnThrTrpHisGlnArgIleAlaLeuLysValGluLeu 409
|||:||||| ||| |||:|||||:|||||:|||||
Qy 1201 CTCGGATGTCAGTTTATTCTCTAAAGGTCGCTCCAAACTTACTCAACCTCCACCTCCT 1260
|||:||||| ||| |||:|||||:|||||:|||||
Db 410 IleGlyCysGln-----IleThrGln----- 416
Qy 1261 CGGAACACGATGAC-----CTCAAAACACTACAGCCCTCCAAAATAGCC 1308
|||:||||| ||| |||:|||||:|||||:|||||
Db 417 -----GlyAsnAspSerLeuValTrpArgLysThrSerGlnSerThrSerValSerThr 434
|||:||||| ||| |||:|||||:|||||:|||||
Qy 1309 AAAGTCGTGCGCCCAAAATTTACCAACCACTCAACCTCGCAGTAGCAATGAATTTCT 1368
|||:||||| ||| |||:|||||:|||||:|||||
Db 435 LysLysGluaspGluThrIleThrArgProIle----- 445
Qy 1369 GCACAGACAGAAACAACTGCCAGTCCTGATATCAGAAATACTACCGTAACCTCCAAAT 1428
|||:||||| ||| |||:|||||:|||||:|||||
Db 446 ---ProSerGluGluThr-----SerThrGlyIleAsnIleThr----- 458
Qy 1429 GTAACCAAGATGAGCGCTGGCGAGTCTTCTCCCTGCGTGCATGCTGCTCCTCACT 1488
|||:||||| ||| |||:|||||:|||||:|||||
Db 459 -----ValAlaIleProLeuValLeuValLeuValPheAlaGlyMet 474
|||:||||| ||| |||:|||||:|||||:|||||
Qy 1489 ACTCTATCTCATATAGTGTGCTTGGCACTGGAGAACAGAAAGAAACAACTGAA 1548
|||:||||| ||| |||:|||||:|||||:|||||
Db 475 GlyIlePheAlaIlePhe-----ArgLysLysLysLysLys 486
Qy 1549 GGCACCTATGACTTACCTTACTTCTGG-----GACCGGGCAGGTGGTGGAAAGGA 1596
|||:||||| ||| |||:|||||:|||||:|||||
Db 487 GlySer-----ProTyrGlySerAlaGluAlaGlnLysThrAspCysTrpLysGln 503
|||:||||| ||| |||:|||||:|||||:|||||
Qy 1597 ATGAAGCAGTTTCTCTCGCAAAAGCAGTGGACCATGAGGAACCCCA---GTTCTGCTAT 1653
|||:||||| ||| |||:|||||:|||||:|||||
Db 504 IleLysTyr-----ProPheAlaArgHis 511
Qy 1654 AGCAGACCGAA-----GTTAATCACTGAGTCCAGAGAGTCACC-----ACA 1698
|||:||||| ||| |||:|||||:|||||:|||||
Db 512 GlnSerAlaGluPheThrIleSerTyrAspAsnGluLysGluMetThrGlnLysLeuAsp 531
Qy 1699 GTGTCGAGGTGAGTCTCTGAGATGATGCTCAGCCACTGTGAGGAAATTTGTTGATCA 1758
|||:||||| ||| |||:|||||:|||||:|||||
Db 532 LeuIleThrSerAspMetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrVal 551
|||:||||| ||| |||:|||||:|||||:|||||
Qy 1759 CTCATCAAGATCTACCTTTAAACCA-----GAAGAGGAAAGAGCAGGCTATGCA 1812
|||:||||| ||| |||:|||||:|||||:|||||
Db 552 ThrArgLysGlySerThrPheArgProMetAspThrAspAlaGluAlaGly---Val 570
|||:||||| ||| |||:|||||:|||||:|||||
Qy 1813 GACCTAGATCCT-----TACAACCTCAGGCGAGGAGGTTTATCATCGCTATGCT 1863
|||:||||| ||| |||:|||||:|||||:|||||
Db 571 SerThrAspAlaGlyGlyHisTyrAspCysProGlnArgAlaGlyArgHisGluTyrAla 590
Qy 1864 GAACCACTCCCAATTTACGGGCTGAGTATGCAACCCCAATCATCATGACATGTCAGGG 1923
|||:||||| ||| |||:|||||:|||||:|||||
Db 591 LeuProLeuAlaProGluProGluProGluTyrAlaThrProIleVal-----GluArg 607
Qy 1924 CACCCCACTTCTAGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAACCA 1983
|||:||||| ||| |||:|||||:|||||:|||||
Db 608 HisValLeuArgAlaHisThrPheSerAlaGlnSerGlyTyrArgValProGlyProGln 627
|||:||||| ||| |||:|||||:|||||:|||||
Qy 1984 CCT-----CCCCCACTA-----GTGGGAACCT 2004
|||:||||| ||| |||:|||||:|||||:|||||
Db 628 ProGlyHisLysHisSerLeuSerSerGlyGlyPheSerProValAlaGlyValGlyAla 647
Qy 2005 TACAATACATCTCTCCAGGACTGACAGCTGCTCTCCAGCCAGGCCCATATGATATACC 2064
|||:||||| ||| |||:|||||:|||||:|||||
Db 648 GlnAspGlyAspTyrGlnArgProHisSerAlaGlnProAlaAspArgGlyTyrAspArg 667
|||:||||| ||| |||:|||||:|||||:|||||
Qy 2065 CCGAAAGCT 2073
|||:||||| ||| |||:|||||:|||||:|||||
Db 668 ProLysAla 670

RESULT 8

ABG01298
ID ABG01298 standard; Protein: 197 AA.

XX AC ABG01298;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #1289.

XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN W0200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT; Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS65485.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics; forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID No 31657; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 197 AA;

Alignment Scores:

Pred. No.:	9,8e-87	Length:	197
Score:	1039.00	Matches:	197
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	26.18%	Indels:	0
DB:	22	Gaps:	0

US-10-060-830-2 (1-2190) x ABG01298 (1-197)

Plate: LLAM12204 row: g column: 15

High quality sequence stop: 665.

FEATURES

source

1. .1065

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:528102"

/clone_lib="NIH_MGC_72"

/tissue_type="melanotic melanoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.

Average insert size 2 kb. Library constructed by Life

Technologies."

BASE COUNT 312 a 280 c 236 g 237 t

ORIGIN

Query Match 69.1%; Score 190; DB 13; Length 1065;

Best Local Similarity 100.0%; Pred. No. 3.8e-90;

Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAACCTCAGTTGTCAGCCCTCCACATCCACTTTCAGGTCAGGGGAGCAACCTCCCC 60

Db 617 CAACCTCAGTTGTCAGCCCTCCACATCCACTTTCAGGTCAGGGGAGCAACCTCCCC 676

QY 61 CACTAGTGGGAACCTTCAATACACTTCTCTCCAGGACTGACAGTGTCTCTCAGCCAGG 120

Db 677 CACTAGTGGGAACCTTCAATACACTTCTCTCCAGGACTGACAGTGTCTCTCAGCCAGG 736

QY 121 CCAGATGATATACCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCCAGAGAAATTGGTGT 180

Db 737 CCAGATGATATACCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCCAGAGAAATTGGTGT 796

QY 181 ACCAGGTGCC 190

Db 797 ACCAGGTGCC 806

RESULT 4

BF037277

LOCUS

DEFINITION 601460996F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3864454 5',

mRNA sequence.

ACCESSION BF037277

VERSION BF037277.1 GI:10745502

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 942)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs@femail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9605 row: p column: 23

High quality sequence stop: 636.

Location/Qualifiers

1. .942

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:3864454"

/clone_lib="NIH_MGC_66"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

FEATURES

source

1. .942

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:3864454"

/clone_lib="NIH_MGC_66"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo c. Average insert size 1.8 kb. Library constructed by Life Technologies."

BASE COUNT 281 a 176 c 232 g 253 t

ORIGIN

Query Match 18.9%; Score 52; DB 12; Length 942;

Best Local Similarity 100.0%; Pred. No. 1.2e-16;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 GGAAGGATGGGGAATGTGATGTTTAAAGAAATCCTTTGAAGATGATGCT 270

Db 1 GGAAGGATGGGGAATGTGATGTTTAAAGAAATCCTTTGAAGATGATGCT 52

RESULT 5

AW326693

LOCUS

DEFINITION 19588 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION AW326693

VERSION AW326693.1 GI:6762614

KEYWORDS EST.

SOURCE cow.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 305)

Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,

Casas, E., Wray, J.E., White, J., Cho, J., Fahnenkrug, S.C., Bennett,

G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,

Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and

Keefe, J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA

libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 20

and -mismatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCTCCAGTCACGACG

Plate: 10 row: B column: 14

Seq primer: ATTTAGGTGACACTATAG.

Location/Qualifiers

1. .305

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 2BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;

Library made from pooled tissue from testis, thymus,

semitendinosus muscle, longissimus muscle, pancreas,

adrenal, and endometrium."

BASE COUNT 80 a 80 c 76 g 59 t

ORIGIN

Query Match 8.4%; Score 23; DB 10; Length 305;

Best Local Similarity 100.0%; Pred. No. 0.29;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 CAGTGTCTCTCAGCCAGGCC 123

Db 53 CAGTGTCTCTCAGCCAGGCC 75

```

RESULT 6
BF543094
LOCUS
DEFINITION
BF543094
ACCESSION
BF543094
VERSION
BF543094.1 GI:11634201
KEYWORDS
EST.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 345)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LLNL (info@image.llnl.gov). IMAGE ID= 1790676
Seq primer: M13 Forward.
FEATURES
source
1..345
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-AFL-aar-a-02-0-UI"
/clone_lib="UI-R-AFL"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AFL
library is a normalized library constructed from 15 dpc
rat atrioventricular (AV) canal. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996. Tissue provided by Jim Lin, Department of Biology,
University of Iowa."
BASE COUNT 89 a 108 c 83 g 65 t
ORIGIN
Query Match 8.4%; Score 23; DB 12; Length 345;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CTTCTCTCCAGGACTGACAGCTG 106
|||||
Db 305 CTTCTCTCCAGGACTGACAGCTG 327

RESULT 7
AI596884
LOCUS
DEFINITION
AI596884
ACCESSION
AI596884
VERSION
AI596884.1 GI:4605932
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 447)
AUTHORS
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE
The WashU-NCI Mouse EST Project 1999
JOURNAL
Unpublished (1999)
COMMENT
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:485015
This read is a RESEQUENCE of a previously sequenced mouse clone
correct orientation)
Seq primer: -40RP from Gibco.
FEATURES
source
1..447
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:808671"
/clone_lib="Soares mouse NbMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCACCAATCTGAAGTGGAGCGCGCGAAGTTTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 114 a 97 c 110 g 126 t
ORIGIN
Query Match 8.4%; Score 23; DB 9; Length 447;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CTTCTCTCCAGGACTGACAGCTG 106
|||||
Db 26 CTTCTCTCCAGGACTGACAGCTG 48

RESULT 8
AA467215
LOCUS
DEFINITION
AA467215
ACCESSION
AA467215
VERSION
AA467215.1 GI:2193355
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 459)
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 447)
AUTHORS
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
TITLE
The WashU-NCI Mouse EST Project 1999
JOURNAL
Unpublished (1999)
COMMENT
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:485015
This read is a RESEQUENCE of a previously sequenced mouse clone
correct orientation)
Seq primer: -40RP from Gibco.
FEATURES
source
1..447
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:808671"
/clone_lib="Soares mouse NbMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCACCAATCTGAAGTGGAGCGCGCGAAGTTTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 114 a 97 c 110 g 126 t
ORIGIN
Query Match 8.4%; Score 23; DB 9; Length 447;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CTTCTCTCCAGGACTGACAGCTG 106
|||||
Db 26 CTTCTCTCCAGGACTGACAGCTG 48

RESULT 8
AA467215
LOCUS
DEFINITION
AA467215
ACCESSION
AA467215
VERSION
AA467215.1 GI:2193355
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 459)
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

```

Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LINL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:485015
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 459.

FEATURES
source
1. .459
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:808671"
/clone_lib="Soares mouse NBMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGAAGTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia) digested with Not I and cloned into the Not
I and Eco RI sites of the modified p7T3 vector. RNA
provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

BASE COUNT 120 a 99 c 113 g 127 t

ORIGIN
Query Match 8.4%; Score 23; DB 9; Length 459;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CTTCTCTCCAGGACTGACAGCTG 106
|||||
Db 26 CTTCTCTCCAGGACTGACAGCTG 48

RESULT 9
BE098908/c
LOCUS
DEFINITION
UI-R-BJ1-atg-b-12-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
UI-R-BJ1-atg-b-12-0-UI 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 479)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

BE098908
LOCUS
DEFINITION
UI-R-BJ1-atg-b-12-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
UI-R-BJ1-atg-b-12-0-UI 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 479)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: msouares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized atrium at 16.5 dpc library cDNA Library Preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source
1. .479
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ1-atg-b-12-0-UI"
/clone_lib="UI-R-BJ1"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
library is a subtracted library derived from the following
tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
For a detailed description of the library from which this
clone was derived, please visit our web site at
rategen.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=UI-R-BJ1
TAG_TISSUE=atrium at 16.5 dpc
TAG_SEQ=CATTC"

BASE COUNT 127 a 105 c 100 g 147 t

ORIGIN
Query Match 8.4%; Score 23; DB 10; Length 479;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CTTCTCTCCAGGACTGACAGCTG 106
|||||
Db 468 CTTCTCTCCAGGACTGACAGCTG 446

RESULT 10
BG802724
LOCUS
DEFINITION
0184-46 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 546)
Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
Nucleic Acids Res. 29 (24), 4983-4993 (2001)
21671825
Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.

FEATURES
source
1. .546
Location/Qualifiers
/organism="Mus musculus"

```

/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps
(Manniat); Cloning Technique: CUA Cloning (Clontech,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACCTCCAGCTCAATTCGAGTG-->. Other
information regarding entire library may be found at
http://pga.swmed.edu/Data/Libraries/microarray_cdna_library
ies.htm"
141 a 133 c 129 g 143 t

BASE COUNT
ORIGIN

Query Match      8.4%; Score 23; DB 12; Length 546;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CTCTCTCCAGGACTGACAGCTG 106
Db 103 CTCTCTCCAGGACTGACAGCTG 125

RESULT 11
LOCUS AK006805
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:1700055P21; endothelial and smooth muscle
cell-derived neuropilin-like molecule, full insert sequence.
ACCESSION AK006805
VERSION AK006805.1 GI:12840068
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
clone_lib:RIKEN full-length enriched mouse cDNA library
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,

```

```

Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bofunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gusticich, S., Hill, D.,
Horrmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marzocchi, L., Mashima, J., Mazzarelli, J., Mombert, P., Nordone, P.,
Ringo, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5 (bases 1 to 1272)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirao, T., Hori, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yamanaka, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adaptor of sequence[5'
GAGAGAGAGAGAGCGCCCAATTAATTCGAGTAATTAATTAATTCGCCCCC 3']. cDNA
was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI, 3'
end: SstI. Host: SOLR.
FEATURES
Location/Qualifiers
1..1272
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="PANTOM.DB:1700055P21"
/db_xref="MGD:MGI:1906364"
/db_xref="taxon:10090"
/clone="1700055P21"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
479..1234
/notice="data source:MGD, source key:MGI:1920629,
evidence:ISS
endothelial and smooth muscle cell-derived neuropilin-like
molecule
putative"
/codon_start=1
CDS

```

```

/dbx_ref="BAB24750.1"
/dbx_ref="GI:12840069"
/dbx_ref="MGD:MG1:1920629"
/translation="MMYVALAAVLPLVLMALTTLILVCWHNRNRKKTGEAYDL
PHMDRAGWKGKMKOLLPAKSDVDEETPVRYSTSEVSHLSAREVTVLOADSAEYAOPL
VGGIVGTWHORSTFKPEGKEAGYADLPDPSMQEYVHAYAEPLVPGPEVATPVM
DMSGPTASVGLPSTSTKTAGTQPHALVGTNTLLSRDSCSSGQAQYDTFAGKSA
AMPSELVQVPOSTQLSGRDEKDFAKELL"
BASE COUNT      283 a   376 c   299 g   314 t
ORIGIN
Query Match      8.4%; Score 23; DB 11; Length 1272;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 CTCTCTCCAGGACTGACAGCTG 106
|||||
Db 1061 CTCTCTCCAGGACTGACAGCTG 1083

RESULT 12
AV204657      298 bp      mRNA      linear      EST 30-OCT-1999
LOCUS
DEFINITION
AV204657      RIKEN full-length enriched, adult male testis Mus musculus
CDNA clone 1700055P21 3', mRNA sequence.
AV204657      GI:6145510
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 298)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Tateno,M., Tomimaga,N., Tsunoda,Y.,
Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gscc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES
source      Location/Qualifiers
1..298
/organism="Mus musculus"
/strain="C57BL/6J"

```

```

/dbx_ref="taxon:10090"
/clone="1700055P21"
/clone_lib="RIKEN full-length enriched, adult male testis"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGAGCGGCCCAATTAATCTCGAGTTAATTAATTAATCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites."
BASE COUNT      74 a   83 c   70 g   71 t
ORIGIN
Query Match      7.6%; Score 21; DB 9; Length 298;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 TATCAGGAGCAGGAGGATG 228
|||||
Db 213 TATCAGGAGCAGGAGGATG 233

RESULT 13
AV205839      324 bp      mRNA      linear      EST 30-OCT-1999
LOCUS
DEFINITION
AV205839      RIKEN full-length enriched, adult male testis Mus musculus
CDNA clone 1700082118 3', mRNA sequence.
AV205839
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 324)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
Suzuki,H., Takahashi,F., Tateno,M., Tomimaga,N., Tsunoda,Y.,
Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gscc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES
source      Location/Qualifiers
1..298
/organism="Mus musculus"
/strain="C57BL/6J"

```


data is from the 3' end
High quality sequence stop: 439.

FEATURES

source

Location/Qualifiers

1. .449
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:808671"
/clone_lib="Soares mouse NbMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5,
TGTTACCAATCTGAGTGGGCGCGGAAAGTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Minoru KO, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldi."

BASE COUNT 139 a 104 c 80 g 126 t
ORIGIN

Query Match 7.6%; Score 21; DB 9; Length 449;

Best Local Similarity 100.0%; Pred. No. 3.5;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 TATCAGGAGCAGGAGGGATG 228

|||||

Db 444 TATCAGGAGCAGGAGGGATG 424

Search completed: January 21, 2003, 14:43:24

Job time : 1386.93 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 12:59:12 ; Search time 35.0787 Seconds
(without alignments)
2404.194 Million cell updates/sec

Title: US-10-060-830-6
Perfect score: 275
Sequence: 1 caacttcagttgtcagccc.....tttgaagatgatgtgttt 275

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/lna/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/lna/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/lna/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/lna/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/lna/FCRUS_COMB.seq.*
6: /cgn2_6/ptodata/2/lna/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	6.5	26385	US-08-961-527-3	Sequence 3, Appli
2	17	6.2	2409	US-09-484-970B-101	Sequence 101, App
3	17	6.2	3083	US-09-484-970B-169	Sequence 169, App
C 4	17	6.2	13011	US-08-791-849A-14	Sequence 14, Appl
C 5	17	6.2	4403765	US-09-103-840A-2	Sequence 2, Appli
6	16	5.8	71	US-09-363-939A-36	Sequence 36, Appl
7	16	5.8	180	US-08-313-075A-41	Sequence 41, Appl
C 8	16	5.8	424	US-08-589-939-8	Sequence 8, Appli
C 9	16	5.8	558	US-08-647-368A-3	Sequence 3, Appli
C 10	16	5.8	659	US-09-328-111-656	Sequence 656, App
C 11	16	5.8	1317	US-07-791-931-1	Sequence 1, Appli
C 12	16	5.8	1376	US-09-443-184-44	Sequence 44, Appl
C 13	16	5.8	1422	US-09-227-357-18	Sequence 18, Appl
C 14	16	5.8	2030	US-09-484-970B-152	Sequence 152, App
C 15	16	5.8	3057	US-09-150-460B-5	Sequence 5, Appli
C 16	16	5.8	3214	US-08-484-105-17	Sequence 17, Appl
17	16	5.8	3214	US-08-484-106-17	Sequence 17, Appl
C 18	16	5.8	4656	US-09-150-460B-4	Sequence 4, Appli
C 19	16	5.8	6614	US-09-150-460B-3	Sequence 3, Appli
C 20	16	5.8	12385	US-09-822-862-3	Sequence 3, Appli
C 21	15	5.5	179	US-09-280-116-97	Sequence 97, Appl
C 22	15	5.5	318	US-08-505-539-4	Sequence 4, Appli
C 23	15	5.5	629	US-09-280-116-228	Sequence 228, App
C 24	15	5.5	790	US-09-378-238-18	Sequence 18, Appl
C 25	15	5.5	920	US-08-299-849B-22	Sequence 22, Appl
C 26	15	5.5	920	US-08-142-368A-22	Sequence 22, Appl
C 27	15	5.5	920	US-08-967-727-22	Sequence 22, Appl

C 28	15	5.5	920	4	US-08-037-230D-22	Sequence 22, Appl
C 29	15	5.5	1128	4	US-09-451-501-22	Sequence 22, Appl
C 30	15	5.5	1128	4	US-09-451-501-26	Sequence 26, Appl
C 31	15	5.5	1167	4	US-09-069-023-19	Sequence 19, Appl
C 32	15	5.5	1668	1	US-08-463-090B-1	Sequence 1, Appli
C 33	15	5.5	1926	2	US-08-464-517-5	Sequence 5, Appli
C 34	15	5.5	1926	3	US-08-463-772-5	Sequence 5, Appli
C 35	15	5.5	1926	5	PCT-US93-05000-5	Sequence 5, Appli
C 36	15	5.5	1962	2	US-08-246-361A-5	Sequence 3, Appli
C 37	15	5.5	2376	1	US-07-912-952-3	Sequence 3, Appli
C 38	15	5.5	2675	2	US-08-883-070-1	Sequence 1, Appli
C 39	15	5.5	2702	4	US-09-766-381-3	Sequence 3, Appli
C 40	15	5.5	2851	4	US-09-535-521-1	Sequence 1, Appli
C 41	15	5.5	2851	4	US-09-535-521-3	Sequence 3, Appli
C 42	15	5.5	3117	3	US-08-909-954-3	Sequence 3, Appli
C 43	15	5.5	3231	4	US-09-134-001C-1200	Sequence 1200, Ap
C 44	15	5.5	3324	2	US-08-820-170A-33	Sequence 33, Appl
C 45	15	5.5	3324	3	US-09-055-699-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-08-961-527-3
; Sequence 3, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/961,527
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26385 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-3

Query Match 6.5%; Score 18; DB 4; Length 26385;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ACTTCAGTTGTCAGCCC 20

Db 22996 ACTTCAGTTGTCAGCCC 23013

RESULT 2

US-09-484-970B-101
 ; Sequence 101, Application US/09484970B
 ; Patent No. 6426186
 ; GENERAL INFORMATION:
 ; APPLICANT: Jones, Karen A.
 ; APPLICANT: Volkmut, Michael G.
 ; APPLICANT: Walker, Michael G.
 ; TITLE OF INVENTION: BONE REMODELING GENES
 ; FILE REFERENCE: PB-0014 US
 ; CURRENT APPLICATION NUMBER: US/09/484,970B
 ; CURRENT FILING DATE: 2000-01-18
 ; NUMBER OF SEQ ID NOS: 172
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 101
 ; LENGTH: 2409
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; OTHER INFORMATION: Incyte ID No. 6426186 404040.2CB1
 ; NAME/KEY: unsure
 ; LOCATION: 11-13, 15-17, 25, 35, 1273, 1281, 1288, 2402
 ; OTHER INFORMATION: a, t, c, g, or other
 US-09-484-970B-101

Query Match 6.2%; Score 17; DB 4; Length 2409;

Best Local Similarity 100.0%; Pred. No. 9.6;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 TGATGTTTTTAAGAAA 252
 |||||
 Db 1312 TGATGTTTTTAAGAAA 1328

RESULT 3

US-09-484-970B-169
 ; Sequence 169, Application US/09484970B
 ; Patent No. 6426186
 ; GENERAL INFORMATION:
 ; APPLICANT: Jones, Karen A.
 ; APPLICANT: Volkmut, Michael G.
 ; APPLICANT: Walker, Michael G.
 ; TITLE OF INVENTION: BONE REMODELING GENES
 ; FILE REFERENCE: PB-0014 US
 ; CURRENT APPLICATION NUMBER: US/09/484,970B
 ; CURRENT FILING DATE: 2000-01-18
 ; NUMBER OF SEQ ID NOS: 172
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 169
 ; LENGTH: 3083
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; OTHER INFORMATION: Incyte ID No. 6426186 237668.3
 US-09-484-970B-169

Query Match 6.2%; Score 17; DB 4; Length 3083;

Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 GCCCAGGCCCATGTA 130
 |||||
 Db 272 GCCCAGGCCCATGTA 288

RESULT 4

US-08-791-849A-14/c
 ; Sequence 14, Application US/08791849A
 ; Patent No. 5914449
 ; GENERAL INFORMATION:

; APPLICANT: Makoto MURASE et al.
 ; TITLE OF INVENTION: Method for Increasing Storage
 ; TITLE OF INVENTION: Lipid Content in Plant Seed
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Ponack
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/791,849A
 ; FILING DATE: January 30, 1997
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION NUMBER:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-8850
 ; TELEFAX:
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 13011 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Rat (Rattus norvegicus)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: Join(3212..3218, 3766..3948, 5917..6008,
 ; LOCATION: 6152..6283, 6418..6604, 6921..7191, 7302..7452, 7665..7817, 7911..8
 ; LOCATION: 9298..9479, 10163..10269)
 ; FEATURE:
 ; NAME/KEY: exon
 ; LOCATION: Join(3194..3218, 3766..3948, 5917..6008,
 ; LOCATION: 6152..6283, 6418..6604, 6921..7191, 7302..7452, 7665..7817, 7911..8
 ; FEATURE:
 ; NAME/KEY: intron
 ; LOCATION: Join(3219..3765, 3949..5916, 6009..6151,
 ; LOCATION: 6284..6417, 6605..6920, 7191..7301, 7453..7664, 7818..7910, 8078..8
 ; LOCATION: 9480..10162)
 US-08-791-849A-14

Query Match 6.2%; Score 17; DB 2; Length 13011;

Best Local Similarity 100.0%; Pred. No. 10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 TCCAGGACTGACAGCTG 106
 |||||
 Db 337 TCCAGGACTGACAGCTG 321

RESULT 5

US-09-103-840A-2/c
 ; Sequence 2, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.

QY 1597 ATGAGCAGTTTCTCTCTCAAGAGCAGTGACCATGAGGAAACCCAGTTCGCTATAGC 1656
Db 1 MetLysGlnPheLeuProAlaLysAlaValAspHisGluThrProValArgTyrSer 20
QY 1657 AGCAGGAGTTAATCACCTGAGTCAAGAGAAGTACCACAGTGTGCGAGGCTGACTCT 1716
Db 21 SerSerGluValAsnHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSer 40
QY 1717 GCAGAGTATGCTCAGCAGCTGTTAGGAGGAATGTTGGTACACTTCAATCAAGATCTACC 1776
Db 41 AlaGluTyrAlaGlnProLeuValGlyGlyValGlyThrLeuHisGlnArgSerThr 60
QY 1777 TTTAAACCAAG 1836
Db 61 PheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerPro 80
QY 1837 GGCAGAGAGTTTATCATGCTATGCTGATGACCACTCCCAATACGGGGCTGAGTATGCA 1896
Db 81 GlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAla 100
QY 1897 ACCCAATCATCATGACATGTGAGGACCCACCACTTTCAGTTGCTGAGCCCTCCACA 1956
Db 101 ThrProIleIleMetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThr 120
QY 1957 TCCACTTTCAAGCTACGGGGAACCACTCCCTCCCTAGTGGAACTTACATFACACTT 2016
Db 121 SerThrPheLysAlaThrGlyAsnGlnProProProLeuValGlyThrTyrAsnThrLeu 140
QY 2017 CTCTCAGGAGTACAGCTCTCTCTCAGCCAGCCAGTATGATACCCGAAAGCTGGG 2076
Db 141 LeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGly 160
QY 2077 AAGCAGGCTACCTGCCAGAGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2136
Db 161 LysProGlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGlu 180
QY 2137 GTATCAGGAGCAGAGAGGATGGGAATGTGATGTTTAAAGAAATCCCTT 2187
Db 181 ValSerGlyAlaGlyArgAspGlyGluCysAspValPheLysGluIleLeu 197
RESULT 9
AAU00630
ID AAU00630 standard; Protein; 539 AA.
XX
AC AAU00630;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #3.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN W0200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX WPI; 2001-290917/30.
DR N-PSDB; AAS00615.
XX
PT Novel nucleic acid encoding human CUB-domain containing protein, useful

PT for drug screening, diagnosis and treatment of physiological disorders
or diseases -
XX
PT
PS Claim 5; Page 29-30; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
domain (an extracellular domain). CUB proteins have been associated with
regulating development, modulating cellular processes and preventing
infectious disease. NHP nucleotide sequences are useful for gene therapy
of physiological disorders or diseases. NHP oligonucleotides are useful
as hybridisation probes for screening libraries and assessing gene
patterns. NHP nucleotide sequences are useful for detecting mutant or
inappropriately expressed NHPs (for example, those proteins associated
with obesity, high blood pressure, connective tissue disorders and
infertility) for the diagnosis of a disease. The polynucleotides may also
be used in screening for drugs effective in the treatment of symptomatic
or phenotypic manifestations of perturbing the normal function of NHP in
the body. Nucleotide constructs encoding NHP products are used to
genetically engineer host cells to express such products in vivo. These
host cells allow for the identification of compounds that bind to NHP
receptors or trigger NHP-mediated pathways.
XX
SQ Sequence 539 AA;

Alignment Scores:
Pred. No.: 4,98e-76 Length: 539
Score: 924.50 Matches: 221
Percent Similarity: 54.34% Conservative: 98
Best Local Similarity: 37.65% Mismatches: 185
Query Match: 23.29% Indels: 83
DB: Gaps: 17

US-10-060-830-2 (1-2190) x AAU00630 (1-539)

QY 13 CTCCTGCTCTTACTTGTCTGCTCTGCTCGAGGAGCTGGAGCCAGCAGGTGAT 72
Db 20 LeuAlaLeuLeuAlaValSerAlaProLeuArgLeuGlnAlaGluLeuGlyAsp 39
QY 73 GGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAACTTACATCAATACCA 132
Db 40 GlyCysGlyHisLeuValThrTyrGlnAspSerGlyThrMetThrSerLysAsnTyrPro 59
QY 133 CAGACCTATCCCAACAGCACCTGTTTGTGATGGAGAGCTCGTGAAGATCGGAGAGA 192
Db 60 GlyThrTyrProAsnHisThrValCysGluLysThrIleThrValProLysGlyLysArg 79
QY 193 GTTCGATCAAAATTTGGTGTGACTTTGACATTCGAGATTCCTGATTCCTCACTTAATAC 252
Db 80 LeuIleLeuArgLeuGlyAspLeuAspIleGlu---SerGlnThrCysAlaSerAspTyr 98
QY 253 TTGAGAATTTATATGGAATTTGGAGTACAGCAACTGAAATAGCAATCTCTGTTCTG 312
Db 99 LeuLeuPheThrSer-----SerSerAspGlnTyrGlyProTyrCysGly--- 113
QY 313 GGGTTGCAAAATGAACCAATTCATTAATGAATCAAAAGCAATGAATCAATCTGTTCTATG 372
Db 114 SerMetThrValProLysGluLeuLeuLeuAsnThrSerGluValThrValArgPheGlu 133
QY 373 AGTGAATTCATGTTTCTGGACGCGGATTTTGGCCCTCAVACTCTGTTATAGATAACAA 432
Db 134 SerGlySerHisIleSerGlyArgGlyPheLeuLeuThrTyrAlaSerSerAspHisPro 153
QY 433 GATCTAATTAATTTTGGACACTGCATCCCAATTTTGGAACTGAGTCTCAGTAAGTAC 492
Db 154 AspLeuIleThrCysLeuGluArgAlaSerHisTyrLeuLysThrGluTyrSerLysPhe 173
QY 493 TGCCAGCTGTTGCTGCT 552
Db 174 CysProAlaGlyCysArgAspValAlaGlyAspIleSerGlyAsnMetValAspGlyTyr 193
QY 553 AGAGATTCCTCGCATTTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 612
Db 194 ArgAspThrSerLeuLeuCysLysAlaAlaIleHisAlaGlyIleIleAlaAspGluLeu 213

```
QY 613 GCGGCCAAATCAGTGTGTAAATAGTAAGGATATCCCTATTATGAAAGTCTCTTGCT 672
    |||||
Db 214 GlyGlyGlnIleSerValLeuGlnArgLysGlyIleSerArgTyrGluGlyIleLeuAla 233
QY 673 AACACGTCACATCTGTGGGGACACATCTATCTACAAGTCTTTTACATTAAGACAAAGT 732
    |||||
Db 234 AsnGlyValLeuSerArgGlySerLeuSerAspLysArgPheLeuPheThrSerAsn 253
QY 733 GGATGTTATGAACACACACACACACACACACACACACACACACACACACACACACAC 792
    |||||
Db 254 GlyCysSerArgSerLeuSerPheGlu-----ProAspGlyGlnIleArgAla 269
QY 793 TCATCTGTCTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 852
    |||||
Db 270 SerSerTrpGlnSerValAsnGluSerGlyAspGlnValHisTrpSerProGlyGln 289
QY 853 CCGAGGCTGAAAAACCTGGACCCCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 903
    |||||
Db 290 AlaArgLeuGlnAspGlnGlyProSerTrpAlaSerGlyAspSerSerAsnHisLys 309
QY 904 ---TACCAGTGGTTACAAATAGATTGAATAAGGAAAGAAATAACAGGCATTATAACC 960
    |||||
Db 310 ProArgGlnTrpLeuGluIleAspLeuGlyGlnLysLysIleThrGlyIleArgThr 329
QY 961 ACTGGATCCACCATGGTGGAGCACATTTACTATGTCTGCTACAGAACTCTGTACAGT 1020
    |||||
Db 330 ThrGlySerThrGlnSerAsnPheAsnPheTyrValLysSerPheValMetAsnPheLys 349
QY 1021 GATGATGGCAGAAATGGACTGTGTACAGACAGCCTGGTGGAGCAGACAGATATTT 1080
    |||||
Db 350 AsnAsnAsnSerLysTrpLysThrTyrLysGlyIleValAsnAsnGluGluValPhe 369
QY 1081 CAAGGAAACAAAGATTCACACAGATGGTGTAACTATTTTCCACCAATATTATGCA 1140
    |||||
Db 370 GlnGlyAsnSerAsnPheArgAspProValGlnAsnAsnPheIleProIleValAla 389
QY 1141 CGTTTATTAGAGTGAATCTTACCCTACCAATGGCAGCAGAAATTTGCCATGAAATGGAGCTG 1200
    |||||
Db 390 ArgTyrValArgValProGlnThrTrpHisGlnArgIleAlaLeuLysValGluLeu 409
QY 1201 CTCGGATGTCAGTTATTCCTAAAGGTCGCTCTCCAAACTTACTCAACCTCCACCTCCT 1260
    |||||
Db 410 IleGlyCysGln-----IleThrGln----- 416
QY 1261 CGGAACACCAATGAC-----CTCAAAACACACTACAGCCCTCCAAAAATAGCC 1308
    |||||
Db 417 -----GlyAsnAspSerLeuValTrpArgLysThrSerGlnSerThrSerValSerThr 434
QY 1309 AAAGTCTGTCGCCCAAAATTTACGCAACACACTACAACTCGCAGTAGCAATGAATTTCT 1368
    |||||
Db 435 LysLysGluAspGluThrIleThrArgProIle----- 445
QY 1369 GCACAGACAGAAACAAACACTGCCAGTCTGATATCAGAAATATACGTTAACTCCAAAT 1428
    |||||
Db 446 ---ProSerGluGluThr-----SerThrGlyIleAsnIleThrThr----- 458
QY 1429 GTAACCAAGATGATGCGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1488
    |||||
Db 459 -----ValAlaIleProLeuValLeuValValLeuValPheAlaGlyMet 474
QY 1489 ACTCTCATTCATATTAAGTGTGTGCTTGGCAGTGGAGAAACAGAAAGAAACAGAA 1548
    |||||
Db 475 GlyIlePheAlaAlaPhe-----ArgLysLysLysLysLys 486
QY 1549 GGCACCTATGACTTACCTTATCTG-----GACCGGCGAGGTGGTGGAAAGGA 1596
    |||||
Db 487 GlySer-----ProTyrGlySerAlaGluAlaGlnLysThrAspCysTrpLysGln 503
QY 1597 ATGAAGCAGTTCTCTCCGCAAGACAGCTGGACCATGAGGAACCCCA---GTCGCTAT 1633
    |||||
Db 504 IleLysTyr-----ProPheAlaArgHis 511
```

```
QY 1654 AGCAGCAGCGAA-----GTTAATCAGTCTGAGTCCAAAGAGAAGTACACACAGTCTGCAG 1707
    |||||
Db 512 GlnSerAlaGluPheThrIleSerTyrAspAsnGlnLysGluMetThrGlnLysLeuAsp 531
QY 1708 CTTGACTCTGCAGAGTATGCT 1728
    |||||
Db 532 LeuIleThrSerAspMetAla 538
RESULT 10
AAU00629
ID AAU00629 standard; Protein; 586 AA.
XX
AC AAU00629;
XX
DT 29-AUG-2001 (first entry)
XX
DE Novel human protein (NHP) sequence #2.
XX
KW Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW obesity; high blood pressure; connective tissue disorder; infertility;
KW NHP-mediated pathway.
XX
OS Homo sapiens.
XX
PN W0200129219-A1.
XX
PD 26-APR-2001.
XX
PF 08-OCT-2000; 2000WO-US28798.
XX
PR 19-OCT-1999; 99US-0160285.
PR 18-FEB-2000; 2000US-0183583.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX
WP1; 2001-290917/30.
N-PSDB; AAS00614.
XX
PT. Novel nucleic acid encoding human CUB-domain containing protein, useful
PT for drug screening, diagnosis and treatment of physiological disorders
PT or diseases -
XX
PS Claim 2; Page 27-28; 33pp; English.
XX
CC The sequence represents a novel human protein (NHP) containing a CUB
CC domain (an extracellular domain). CUB proteins have been associated with
CC regulating development, modulating cellular processes and preventing
CC infectious disease. NHP nucleotide sequences are useful for gene therapy
CC of physiological disorders or diseases. NHP oligonucleotides are useful
CC as hybridisation probes for screening libraries and assessing gene
CC patterns. NHP nucleotide sequences are useful for detecting mutant or
CC inappropriately expressed NHPs (for example, those proteins associated
CC with obesity, high blood pressure, connective tissue disorders and
CC infertility) for the diagnosis of a disease. The polynucleotides may also
CC be used in screening for drugs effective in the treatment of symptomatic
CC or phenotypic manifestations of perturbing the normal function of NHP in
CC the body. Nucleotide constructs encoding NHP products are used to
CC genetically engineer host cells to express such products in vivo. These
CC host cells allow for the identification of compounds that bind to NHP
CC receptors or trigger NHP-mediated pathways.
XX
SQ Sequence 586 AA;
```

```
Alignment Scores:
Pred. No.: 5,15e-76 Length: 586
Score: 924.50 Matches: 221
Percent Similarity: 54.34% Conservative: 98
Best Local Similarity: 37.65% Mismatches: 185
Query Match: 22.29% Indels: 83
DB: Gaps: 17
```


XX (LEXI-) LEXICON GENETICS INC.
 PA Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
 XX WPI; 2001-290917/30.
 XX DR N-PSDB; AAS00613.
 XX Novel nucleic acid encoding human CUB-domain containing protein, useful
 PT for drug screening, diagnosis and treatment of physiological disorders
 PT or diseases -
 XX
 XX Claim 4; Page 26; 33pp; English.
 XX
 CC The sequence represents a novel human protein (NHP) containing a CUB
 CC domain (an extracellular domain). CUB proteins have been associated with
 CC regulating development, modulating cellular processes and preventing
 CC infectious disease. NHP nucleotide sequences are useful for gene therapy
 CC of physiological disorders or diseases. NHP oligonucleotides are useful
 CC as hybridisation probes for screening libraries and assessing gene
 CC patterns. NHP nucleotide sequences are useful for detecting mutant or
 CC inappropriately expressed NHPs (for example, those proteins associated
 CC with obesity, high blood pressure, connective tissue disorders and
 CC infertility) for the diagnosis of a disease. The polynucleotides may also
 CC be used in screening for drugs effective in the treatment of symptomatic
 CC or phenotypic manifestations of perturbing the normal function of NHP in
 CC the body. Nucleotide constructs encoding NHP products are used to
 CC genetically engineer host cells to express such products in vivo. These
 CC host cells allow for the identification of compounds that bind to NHP
 CC receptors or trigger NHP-mediated pathways.
 XX

SQ Sequence 487 AA;

Alignment Scores:

Pred. No.: 3.67e-70 Length: 487
 Score: 860.50 Matches: 206
 Percent Similarity: 54.33% Conservative: 95
 Best Local Similarity: 37.18% Mismatches: 170
 Query Match: 21.68% Indels: 83
 DB: 22 Gaps: 17

US-10-060-830-2 (1-2190) x AAU00628 (1-487)

QY 112 CTTACATCAATAAATACCCACAGACCTATCCACAGACCTGTTGTGATGGAGATC 171
 DB 1 MetThrSerLysAsnTyrProGlyThrTyrProAsnHisThrValCysGluLysThrIle 20
 QY 172 CGTGTAAAGATGGGAGAGAGTTGCGCATCAAAATTTGGTGACTTTGACATTGAAGATTCT 231
 DB 21 ThrValProLysGlyLysArgLeuIleLeuArgLeuGlyAspLeuAspIleGlu---Ser 39
 QY 232 GATTCTTGTCTACTTAAATTAATGAGATTATTAATGGAATGGAGTCAGCAGAACTGAA 291
 DB 40 GlnThrCysAlaSerAspTyrLeuLeuPheThrSer-----SerSerAspGln 55
 QY 292 ATAGCAATACTGGTCTGGGTGCTGGGTGCAATGAACCATTCATTAATGAATCAAAAGGCAT 351
 DB 56 TyrGlyProTyrCysGly---SerMetThrValProLysGluLeuLeuAsnThrSer 74
 QY 352 GAAATACATGCTCTCATGATGGGAATCCATGTTCTGGACGGGATTTTGGCCCTCA 411
 DB 75 GluValThrValArgPheGluSerGlySerHisIleSerGlyArgGlyPheLeuLeuThr 94
 QY 412 TACTCTGTTATAGATAAACAAGATCTAATATTGTTGGACACTGCATCAATTTTTG 471
 DB 95 TyrAlaSerSerHisProAspLeuIleThrCysLeuLeuGluAlaSerHisTyrLeu 114
 QY 472 GAACCTGAGTTCAGTAAGTACTGCCACTGGTGTCTGCTCTCTTTTGTCTGAGATATCT 531
 DB 115 LysThrGluTyrSerLysPheCysProAlaGlyCysArgAspValAlaGlyAspIleSer 134
 QY 532 GGAACAATTCCTCATGATATAGATATCTCGCATTCATGTCATGGCTGGTGTGCATCA 591

DB 135 GlyAsnMetValAspGlyTyrArgAspThrSerLeuLeuCysLysAlaAlaIleHisAla 154
 QY 592 GGAGTACTGTCAACACACGTTGGCGGCGCAATAGTGTGTTAATAGTAAGGTATATCCC 651
 DB 155 GlyIleIleAlaAspGluLeuGlyGlyGlnIleSerValLeuGlnArgLysGlyIleSer 174
 QY 652 TATTATGAAAGATCTTTGGGTAAACAACGTCACATCTGTGTGGTGGACACATCTATCTACAAGT 711
 DB 175 ArgTyrGluGlyIleLeuAlaAsnGlyValLeuSerArgAspGlySerLeuSerAspLys 194
 QY 712 CTTTTCATATTAAGACAAGTGGATGTTATGGAACACTGGGGATGGATGCTGTGTGTGATC 771
 DB 195 ArgPheLeuPheThrSerAsnGlyCysSerArgSerLeuSerPheGlu----- 210
 QY 772 CGGATCCTCAATAACAGCATCATCTGTGTGGTGGATGGATGACACACAGGCAAGAG 831
 DB 211 ProAspGlyGlnIleArgAlaSerSerTrpGlnSerValAsnGluSerGlyAspGln 230
 QY 832 AACAGTTGGAAACCCAAAAAGCCAGGCTGAAACCACTGGACCGCTTGGGCTGCTTTT 891
 DB 231 ValHisTrpSerProGlyGlnAlaArgLeuGlnAspGlnGlyProSerTrpAlaSerGly 250
 QY 892 CCCACTGATGAA-----TACAGTGGTTACAATAGATTGTAATGAAGAAAG 939
 DB 251 AspSerSerAsnAsnHisLysProArgGluTrpLeuGluIleAspLeuGlyLysLys 270
 QY 940 AAAATAACAGCATTATACCACTGGATCCACCATGGTGGAGCAACAATTACTATGTCT 999
 DB 271 LysIleThrGlyIleArgThrGlySerThrGlnSerAsnPheAsnPheTyrValLys 290
 QY 1000 GCCTACAGATCCTGTACAGTGCATGGGAGAAATGGACTGTGTACAGACGCGCTGGT 1059
 DB 291 SerPheValMetAsnPheLysAsnAsnSerLysTrpLysThrTyrLysGlyIleVal 310
 QY 1060 GTGGAGCAAGATAAGATATTTCAAGGAACAAGATATATCACCAGATGTGGTAAATAC 1119
 DB 311 AsnAsnGluGlyLysValPheGlnGlyAsnSerAsnPheArgAspProValGlnAsnAsn 330
 QY 1120 TTTTGGCCCAATTAATGACGTTTATTAGATGATCTTACCAATGGCAGCAGAAA 1179
 DB 331 PheIleProIleValAlaArgTyrValArgValAlaProGlnThrTrpHisGlnArg 350
 QY 1180 ATTGCCATGAAATGGAGCTGCTCGATGTGAGTTTATTCTTAAAGTCGCTCTCCAAA 1239
 DB 351 IleAlaLeuLysValGluIleGlyCysGln----- 361
 QY 1240 CTTACTCAACCTCCACCTCCTCGGAACAGCAATGAC-----CTCAAAACACT 1287
 DB 362 IleThrGln-----GlyAsnAspSerLeuValTrpArgLysThrSer 375
 QY 1288 ACAGCCCTCCAAAATAAGCAAGGTGCTGCCCAAAATTTACGCAACCACTACAACCT 1347
 DB 376 GlnSerThrSerValSerThrLysLysGluAspGluThrIleThrArgProIle----- 393
 QY 1348 CGCAGTAGCAATGAATTCCTGCACAGACAGAAACAACCTGCCAGCTCCTCATATCAGA 1407
 DB 394 -----ProSerGluGluThr-----SerThrGlyIleAsn 403
 QY 1408 AATACCTACCTAATCCAAATGTACCAAAAGATGTAGCGCTGGCTGCACTTCTGTCCT 1467
 DB 404 IleThrThr-----ValAlaIleProLeuValLeuVal 415
 QY 1468 GTGTGTGTCATGGTCTCACTACTCTCATCTCATATTCATATTAGTGTGTGTGGCAGCTGGAGA 1527
 DB 416 ValLeuValPheAlaGlyMetGlyIlePheAlaAlaPhe----- 428
 QY 1528 AACGAAGAAGAAAACTGAAGGCACCTATGACTTACCTTACTGG-----GAC 1575
 DB 429 ---ArgLysLysLysLysLysGlySer-----ProTyrGlySerAlaGluAlaGln 444
 QY 1576 CGGCAGGTTGTGGAAAGGAATGAAGCAGTTTCTCTCTGCAAAAGACGCTGGACCATGAG 1635
 DB 445 LysThrAspCysTrpLysGlnIleLysTyr----- 454

QY	1636	GAAACCCCA---GTTTCGTATAGCAGCGA-----GTTAATCACCTGAGTCCAAGA	1686
		::: :::	::: :::
Db	455	-----ProPheAlaArgHisGlnSerAlaGluPhetThrIleSerTyRAspAsnGluLys	472
		:::	::: :::
QY	1687	GAAGTCACCACAGTGCTCGAGGCTGACTGCCTGCACAGTAGTGCT	1728
		:::	::: :::
Db	473	GlumethrGlnLysLeuAspLeulleThrSerAspMetAla	486
		:::	::: :::
RESULT 12			
AAB19126	ID	AAB19126 standard; Protein; 503 AA.	
XX	AC	AAB19126;	
XX	DT	19-FEB-2001 (first entry)	
DE	XX	Polypeptide isolated from lymph node stromal cells of fsn +/- mice.	
XX	XX	Lymph node stromal cell; fsn +/- mice; inflammatory disorder;	
KW	KW	immune system disorder; cancer; viral disorder; HIV infection;	
KW	KW	blood vessel growth; tumour necrosis factor disorder; arthritis;	
KW	KW	inflammatory bowel disease; fibroblast growth factor-mediated disorder	
KW	KW	cardiac failure.	
OS	XX	Mus sp.	
OX	XX	WO200058463-A1.	
PN	XX	05-OCT-2000.	
PD	XX		
PF	XX	18-FEB-2000; 2000WO-NZ00015.	
PP	XX	25-MAR-1999; 99US-02762586.	
PR	XX	26-AUG-1999; 99US-0383586.	
PS	XX	(GENE-) GENESIS RES & DEV CORP LTD.	
PA	XX	Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;	
PI	XX	Murison JG;	
PI	XX	WPI; 2000-664924/64.	
DR	XX	N-PSDB; AAA96736.	
XX	XX		
PT	PT	Polypeptide expressed in mammalian fsn +/- lymph node stromal cells,	
PT	PT	useful for modulating growth of blood cells, for treating inflammatory	
PT	PT	and tumour necrosis factor-mediated disorders, cancer and viral	
PT	PT	disorders	
XX	XX		
PS	XX	Claim 1; Page 68-69; 75pp; English.	
XX	XX	The present sequence represents a polypeptide sequence which is	
CC	CC	isolated from lymph node stromal cells of fsn +/- mice. The	
CC	CC	polynucleotides and their polypeptides are useful for treating an	
CC	CC	inflammatory disorder, disorder of immune system and cancer selected	
CC	CC	from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a	
CC	CC	viral disorder, in particular HIV infection and for modulating the	
CC	CC	growth of blood vessels. The polypeptides are useful for treating a	
CC	CC	tumour necrosis factor (TNF) mediated disorder, such as those selected	
CC	CC	from arthritis, inflammatory bowel disease and cardiac failure and a	
CC	CC	fibroblast growth factor-mediated disorder. It is also useful in assay	
CC	CC	to determine biological activity, to raise antibodies, to isolate	
CC	CC	corresponding ligands or receptors, to quantify levels of protein or	
CC	CC	cognate corresponding ligand or receptors, as anti-inflammatory agents,	
CC	CC	and in compositions for the treatment of skin, connective tissue and	
CC	CC	immune system diseases. The polynucleotide is useful as marker for	
CC	CC	tissue, as a chromosome marker or tags in the identification of a	
CC	CC	genetic disorder.	
XX	XX		
SQ	Sequence	503 AA;	
Alignment Scores:			
Pred. No.:		7,39e-44	Length: 503

Score:	574.50	Matches:	184
Percent Similarity:	35.83%	Conservative:	84
Best Local Similarity:	24.60%	Mismatches:	179
Query Match:	14.47%	Indels:	301
DB:	21	Gaps:	21
US-10-060-830-2 (1-2190) x AAB19126 (1-503)			
Qy	4	CCTGTGTTCTCTGCTCTACTTGTCTGCTCTGCTGAGGACGCTGGAGCCGAC	63
Db	8	ProSerValLeuAlaLeuLeuPheAlaValCysAlaProLeuArgLeuGlnAlaGluGlu	27
Qy	64	CAAGGTGATGGATGGACACACTGTAAGCCCTGAGAGTGGAAACCTTACATCCATA	123
Db	28	LeuGlyAspGlyCysGlyHisIleValThrSerGlnAspSerGlyThrMetThrSerLys	47
Qy	124	AACATACCACAGACCTATCCACACAGACACTGTTGTGAATGGGAGATCCGTGTAAAGATG	193
Db	48	AsnFyrProGlyThrFyrProAsnFyrThrValCysGluLysIleIleThrValProLys	67
Qy	184	GGAGAGAGTTCGCATCAAAATGGTGACTTTGACATTTGAAGATTTCTTCTTCAC	243
Db	68	GlyLysArgLeuIleLeuArgLeuGlyAspLeuAsnIleGlu---SerLysThrCysAla	86
Qy	244	TTTAATTACTTGAAATTTATATAGTAATGGAGTGCAGACAGACTGAATAGGCAATAC	303
Db	87	SerAspFyrLeuLeuPheSerSer-----AlaThrAspGlnTyrGlyProTyr	102
Qy	304	TGTGCTGGGTTGCAATGAACCATTTCAATTAATCAAAAGGCAATGAATACACATTG	363
Db	103	CysGly---SerTrpAlaValProLysGluLeuArgLeuAsnSerAsnGluValThrVal	121
Qy	364	CTGTTTCATGATGAATCCATGTTTCTGGACGCGGATTTTGGCTCATCTCTGTATA	423
Db	122	LeuPheLysSerGlySerHisIleSerGlyArgGlyPheLeuLeuThrTyrAlaSerSer	141
Qy	424	GATAACAAGATCPAATTACTGTTTGGACACTGCATCCAAATTTTGGAACTGAGTTC	483
Db	142	AspHisProAspLeuIleThrCysLeuGluArgGlySerHisTyrPheGluGluLysTyr	161
Qy	484	AGTAAGTACTGCCAGCTGGTTGCTGCTCTTGGCTCAGATATCTGGAAACAATTCCT	543
Db	162	SerLysPheCysProAlaGlyCysArgAspIleAlaArgAspIleSerGlyAsnThrLys	181
Qy	544	CATGGATATAGAGATTCTCGCCATTCTGCATGCTGGTGTGCATGCAGAGAGTAGTGCA	603
Db	182	AspGlyTyrArgAspThrSerLeuLeuCysLysAlaAlaIleHisAlaGlyIleIleThr	201
Qy	604	AACACGTTGGCGGCCAAAATCAGTGTGTTAATTAGTAAGGTATTCCTTATATGAAAGT	663
Db	202	AspGluLeuGlyGlyHisIleAsnLeuLeuGlnSerLysGlyIleSerHisTyrGluGly	221
Qy	664	TCTTTGGCTAACACGTCACATCTGTGGTGGGACACTTATCTACAAGTCTTTTTACATT	723
Db	222	LeuLeuAlaAsnGlyValLeuSerArgHisGlySerLeuSerGluLysArgPheLeuPhe	241
Qy	724	AAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGGTGTGATCGCGGATCCTCA	783
Db	242	ThrThr-----	243
Qy	784	ATAACAGCATCATCTGCTGGATGGAGTGCACACACAGGGCAAGAACAGTGGAA	843
Db	243	-----	243
Qy	844	CCCAAAAAGCCAGGCTGAAAAAACCTGGAGCCGCTTGGCCTGCTTTTGGCCATGTAA	903
Db	243	-----	243
Qy	904	TACCAGTGGTTACAAATAGATTGGAATAAGGAAAGAAATAACAGGCATTATAACCACT	963
Db	243	-----	243
Qy	964	GGATCCACCATGTTGGGACACAAATTAATGTGTCCTGCCTACAGATTCCTGTACAGTGT	102

Db 243 ----- 243
QY 1024 GATGGCAGAAATGGACTGTGTACAGAGAGCCGTGTGTGGAGCAAGATAAGATATTTCAA 1083
Db 244 ----- 246
QY 1084 GGAAACAAAGATTATCACCAGGATGTGCGTAATAAATCTTTTGGCCACCAATTATTGCACGT 1143
Db 246 ----- 246
QY 1144 TTTATTAGAGTATCTTACCAATGCGAGAGAAATGCGATGAAATGGAGCTGCTC 1203
Db 246 ----- 246
QY 1204 GGATGTCAGTTTATTCTTAAGGTGCTCTCCAAACTTACTCAACCTCCACCTCCTCGG 1263
Db 246 ----- 246
QY 1264 AACAGCAATGACCTCAAAACACTACAGCCCTCCAAAATAGCAAGGTGTCGCCCA 1323
Db 246 ----- 246
QY 1324 AAATTTACGCAACCACTACAACTCGCAGTAGCAATGAATTTCTGCACAGACAGAACAA 1383
Db 246 ----- 246
QY 1384 ACAACTGCCAGTCTGATATCAGAAATACTACCTGAATCTCAATGTAAACCAAGATGTA 1443
Db 247 ----- 251
QY 1444 CGCTGGCTGAGTCTTGTCCCTGTGCTGCTGATGTCPCACTACTCTCTCATCTCATATA 1503
Db 252 AlalleProSerValle ----- PhelealeuLeuLeuLeuThrGlyMetGlylePhe 269
QY 1504 TTAGTGTGTGTGCTGAGTGGAGAACAAAGAAAACTGAAGGCACCTATGACTTA 1563
Db 270 AlalleCys ----- ArgLysArgLysLysGlyAsn ----- 280
QY 1564 CTTTAC ----- TGGACGGGCGAGTGGTGGGAAGGANTGAACGAGTTCCTT 1611
Db 281 ProTyValSerAlaAspAlaGlnLysThrGlyCysThrLysGlnIleLysTyr ----- 298
QY 1612 CTGCAAAAGCAGTGGACCATGAGGAACCCCA --- GTTGCCTATAGCAGCAGCAA --- 1665
Db 299 ----- ProPhealaArgHisGlnSerThrGluPhe 308
QY 1666 --- GTTAATCACTGAGTCCAGAGAGTCACTAC --- ACAGTGTGCTGAGGCTGAC 1713
Db 309 ThrIleSerTyAspAsnGluLysGluMetThrGlnLysLeuAspLeuIleThrSerAsp 328
QY 1714 TCTGCAGAGTATGCTCAGCACTGTGAGAGAAATTTGTTGCTACACTTCATCAAGATCT 1773
Db 329 MetAlaAspTyGlnGlnProLeuMetIleGlyThrGlyThrValAlaArgLysGlySer 348
QY 1774 ACCTTTAAACCA ----- CAAGAA ----- GGAAGAGACGAGC 1806
Db 349 ThrPheArgProMetAspThrAspThrGluGluValArgValAsnThrGluAlaSerGly 368
QY 1807 TATGCAGACCTAGATCCTTACAACCTACAGGCGAGGAAGTTTATCATGCTCATGCTCAA 1866
Db 369 His ----- TyrAspCysProHisArgProGlyArgHisGluTyAlaLeu 383
QY 1867 CCACCTCCCAATACGGGCTGTGATGATCAACCCCAATCATC ----- 1908
Db 384 ProLeuThrHisSerGluProGluTyAlaThrProIleValGluArgHisLeuLeuArg 403
QY 1909 ----- ATGACATGTCAGGCGACCCCAACTTCACTGTTGGTCAGCCCTCCACA 1956
Db 404 AlaHisThrPheSerThrGlnSerGlyTyArgValProGly --- ProArgProThrHis 422
QY 1957 TCCACTTTCAAGCTACGGGGAACCACTCCCTCCTAGTGGGAACCTTACATACTACTT 2016

Db 423 GluHisSerHisSerSerGlyGlyPheProProAlaThrGlyAlaThrGlnValGluSer 442
QY 2017 CTCTCAGGAGTACAGCTGCTCTCAGCCAGCCAGTATGATACCCCGAAGCTGGG 2076
Db 443 TyrGlnArgProAlaSerProLysProValGlyGlyGlyTyAspLysPro --- AlaAla 461
QY 2077 AAGCCAGTCTACTCTGCCCGCAGCAATTTGTTGACAGGTGTCACAGACGACACAGAA 2136
Db 462 SerSerPheLeuAspSerArgAsp ----- ProAlaSerGlnSerGln 475
QY 2137 GTATCAGGACGAGGAGGATGGG 2160
Db 476 MetThrSerGlyGlyAspGly 483
RESULT 13
AAE22715
ID AAE22715 standard; Protein; 398 AA.
XX
AC AAE22715;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human neuropilin-Hy1 protein.
XX
KW Human; neuropilin-like polypeptide; neuropilin-Hy1; neovascularisation;
KW neurodegenerative disease; Alzheimer's disease; learning; angiogenesis;
KW thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory;
KW ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide;
KW wound healing; tissue repair; Parkinson's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord;
KW cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis;
KW human immunodeficiency virus; HIV; autoimmune disorder; dermatological;
KW systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer;
KW autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease;
KW myasthenia gravis; autoimmune inflammatory eye disease; gene therapy;
KW neurotropic; neuroprotective; vulnary; anticonvulsant; antiparasitic;
KW cerebroprotective; tranquiliser; virucide; antibacterial; cytostatic;
KW immunosuppressive; chromosome 6q21.
XX
OS Homo sapiens.
XX
PN WO200222815-A1.
XX
PD 21-MAR-2002.
PF 12-SEP-2001; 2001WO-US28488.
XX
PR 11-SEP-2000; 2000US-0659671.
PR 06-SEP-2001; 2001US-317902P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT;
XX
DR WPI: 2002-393966/42.
DR N-PSDB: AAD35992.
XX
PT Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides
PT useful for treating neurodegenerative diseases e.g. Alzheimer's
PT disease, and for diagnosing and mapping genetic neuronal defects
XX
PS Claim 3; Page 123-125; 152pp; English.
XX
CC The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and
CC neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-
CC like polypeptides and polynucleotides are useful in modulating neuronal
CC growth regenerative capacity, treating neurodegenerative diseases,
CC diagnosing and mapping genetic neuronal defects and degenerative diseases
CC like Alzheimer's disease and for treating learning and memory disorders.
CC They are also useful for inducing angiogenesis, neovascularisation, as
CC well as organ growth and development e.g. heart and other tissues.
CC Antagonists of neuropilin-like polypeptides are useful for treating
CC cancers and other malignant diseases. Neuropilin is used to treat

RA Fujisawa H.;
RT "Expression of a cell adhesion molecule, neuropilin, in the
RL developing chick nervous system.";
RL Dev. Biol. 170:207-222(1995).
CC -1- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
CC CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
CC CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
CC SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
CC SIMILARITY). SEEMS TO HAVE CALCIUM-INDEPENDENT CELL ADHESION
CC PROPERTIES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: DEVELOPING NERVOUS SYSTEM; OPTIC TECTUM
CC (LAYERS D AND E OF SGFS), AMACRINE CELLS OF RETINA, NEURITES OF
CC DORSAL ROOT GANGLIA. ALSO EXPRESSED IN NONNEURONAL CELLS, E.G.
CC BLOOD VESSELS IN THE ENTIRE EMBRYO.
CC -1- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D45416; BAA08256.1; -
CC HSP; P12259; 1C2T.
CC InterPro; IPR000859; CUB_domain.
CC InterPro; IPR000421; FA58_C.
CC InterPro; IPR000998; MAM_domain.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00629; MAM; 1.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC PRINTS; PR00020; MAMDOMAIN.
CC SMART; SM00042; CUB; 2.
CC SMART; SM00231; FA58C; 2.
CC SMART; SM00137; MAM; 1.
CC SMART; PS01180; CUB; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PROSITE; PS00740; MAM_1; 1.
CC PROSITE; PS00060; MAM_2; 1.
CC Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Cell adhesion.
FT SIGNAL 1 18
FT CHAIN 19 914
FT DOMAIN 20 847
FT TRANSMEM 848 870
FT DOMAIN 871 914
FT DOMAIN 25 139
FT DOMAIN 145 263
FT DOMAIN 273 422
FT DOMAIN 429 581
FT DOMAIN 636 801
FT DOMAIN 235 52
FT DISULFID 80 102
FT DISULFID 145 171
FT DISULFID 204 226
FT DISULFID 273 422
FT DISULFID 429 581
SQ SEQUENCE 914 AA; 102480 MW; DD2E6D6F0CB68C CRC64;

Alignment Scores:

Pred. No.: 1.26e-19 Length: 914
Score: 384.50 Matches: 131
Percent Similarity: 40.44% Conservative: 72
Best Local Similarity: 26.10% Mismatches: 200
Query Match: 9.28% Indels: 99
Db: 1 Gaps: 19

US-10-060-830-1 (1-2280) x NRPI_CHICK (1-914)
QY 133 GCCAGAGCAAGCTGATGGATGGGACACACTGTACTAGCCCTGAGAGTGGAGACCTTACA 192
DB 19 AlaLeuArgSerAspLysCysGlyAspThrIleLysIleLeuSerProGlyTyrLeuThr 38
QY 193 TCCATAAACTACCACAGACCTATCCCAACAGCACTGTTTGTGAATGGGAGATCGGTGA 252
DB 39 SerProGlyTyrProGlnSerTyrHisProSerGlnLysCysGluTrpLeuIleGlnAla 58
QY 253 ---AAGATGGGAGAGAGTTCGCATCAAAATTTGGT---GACTTTGACATTAAGACTTCT 306
DB 59 ProGluProTyrGlnArgIleMetIleAsnProHisPheAspLeuGluAspArg 78
QY 307 GATCTCTGCTCACTTAATTAATGAGATTTATATGGAATTTGAGTGGAGTCACTGAA 366
DB 79 Asp---CysLysTyrAspTyrValGluValIleAspGlyAspAsnAlaGluGlyArgLeu 97
QY 367 ATAGGCAAACTACTGTGGTCTGGGTTGCAATGAACCATTCATTAATCAAAAGGCAAT 426
DB 98 TrpGlyLysTyrCys-----GlyLysIleAlaProProProLeuValSerSerGlyPro 115
QY 427 GAAATCACTGCTGTTTCATGAGTGAATCCATGTTTCTGGACGCGGATTTTGGCCCTCA 486
DB 116 TyrLeuPheIleLysPheValSerAspTyrGluThrHisGlyAlaGlyPheSerIleArg 135
QY 487 TACTCTGTTATAGATAAA-----
DB 136 TyrGluValPheLysArgGlyProGluCysSerArgAsnPheThrSerSerSerGlyMet 155
QY 505 -----CAAGATCTAATTACTTGT-----
DB 156 IleLysSerProGlyPheProGluLysTyrProAsnSerLeuGluCysThrTyrIleIle 175
QY 523 -----TTGGACACTGCATCCCAATTTTGGAACTGAG 555
DB 176 PheAlaProLysMetSerGluIleIleLeuGluPheGluSerPheGluLeuGluProAsp 195
QY 556 TTCAGTAAGTACTGCCAGCTGGTGTCTGCTTCCTTTGCTGAGATATCTGGAACAATT 615
DB 196 SerAsn-----ThrProGlyGlyAlaPheCysArgTyrAspArgLeuGlu-----Ile 211
QY 616 CCTCATGATATAGATATCTCGCCA-----TTGTGCATGCTGGTGTGCAT 663
DB 212 TrpAspGlyPheProAspValGlyProHisIleGlyArgTyrCysGlyGlnAsnAsnPro 231
QY 664 GCAGGAGTAGTGTCAACACACAGTGGCGGCCCAATAGTGTGTATTAATAGTAAAGTATT 723
DB 232 GlyArgValArgSerSerThr-----GlyIleLeuSerMetValPhe-----
QY 724 CCTATTATGAAAGTCTTTGGCTAACACGTCACATCTGTGGTGGGACACTTATCTACA 783
DB 246 ---TyrThrAspSerAlaIleAlaLysGluGlyPheSerAlaAsnTyrSerValSerGln 264
QY 784 AGTCTTTTACATTTAAGCAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTG 843
DB 265 SerSerValSer---GluAspPheGlnCysMetGluProLeuGlyMetGluSerGlyGlu 283
QY 844 ATCCGGGATCTCTCAATAACACGATCATCTGTGCTGGTGGAGTGGACTGACCACACAGG 903
DB 284 IleHisSerAspGlnIleThrValSerSer-----Gln 294
QY 904 GAGAACAGT---TGGAAACCCCAAAAGACGAGCTGAAACAAACCTGGACCGCTGGGCT 960
DB 295 TyrSerAlaIleTyrSerSerGluArgSerArgLeuAsnTyrProGluAsnGlyTyrThr 314
QY 961 GCTTTTGGCACTGATGAATACCACTGGTGTGTACAAATAGATTTGAATAAGAAAGAAATA 1020
DB 315 ProGlyGluAspSerValArgGluTrpIleGlnValAspLeuGlyLeuLeuArgPheVal 334
QY 1021 ACAGGCATTATAACCACTGGA-----TCCACCATGGTGGAGCACAAATTAATGTGTCT 1074
DB 335 SerGlyIleGlyThrGlnGlyAlaIleSerLysGluThrLysLysGluTyrTyrLeuLys 354


```
Db 81 Glu---CysLysTyrAspTyrValGluValIleAspGlyAspAsnAlaAsnGlyGlnLeu 99
QY 367 ATAGGCAAAATAGTGTGCTGGGTGCAAAATGCAACATTC---ATTGAATCAAAAGGC 423
Db 100 LeuGlyLysTyrCysGly-----LysIleAlaProSerProLeuValSerThrGly 116
QY 424 AATGAATACATGCTGTTCATGAGTGGATCCATGCTTTCTGACGCGGATTTTGGCC 483
Db 117 ProSerIlePheIleArgPheValSerAspTyrGluThrProGlyAlaGlyPheSerIle 136
QY 484 TCATACCTCTGTATAGATAAACAAGACTAATTAATCTGTTGGACACTGCATCAATTT 543
Db 137 ArgTyrGluValPheLys----- 142
QY 544 TTGAACCTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTGCTTCTCTTGTGCTGAGATA 603
Db 143 ThrGlyProGlu-----CysSerArgAsnPheThrSerSer 154
QY 604 TCTGGAACAATT-----CCTCATGGATATAGATTC-----TCG 639
Db 155 AsnGlyValIleLysSerProLysTyrProGluLysTyrProAsnAlaLeuGluCysThr 174
QY 640 CCATGTGCATGCTGTGTCATGCAGGAGTAGTG----- 675
Db 175 TyrIleIlePheAlaProLysMetGlnGluIleValLeuGluPheGluSerPheGluLeu 194
QY 676 -----TCAACACAGTTGGGGGCGCAATCAGT-----GTTGTAATTAGT 714
Db 195 GluAlaAspSerAsnAlaProGlyGlnThrCysArgTyrAspTyrLeuGlyIleTyr 214
QY 715 AAGGTATTCC-----TATTATGAACTTCTTGGCTAAC 750
Db 215 AspGlyPheProGlyValGlyProHisIleGlyArgTyrCysGlyGlnAsnThrProGly 234
QY 751 AACGTACATCTGTGGGGGACCTTCTACAAAGTCTTTTACATTAAGACAAGTGA 810
Db 235 ArgValArgSerPheThrGlyIleLeuSerMetIlePheHisThrAspSerAlaIleAla 254
QY 810 ----- 810
Db 255 LysGluGlyPhePheAlaAsnPheSerValValGlnSerAsnThrAspGluAspPheGln 274
QY 811 TGTATGGAACACATGGGATGGAGTCTGGGTGATCGCGGATCCTCAATAACAGCATCA 870
Db 275 CysLysGluAlaLeuGlyMetGluSerGlyGluIleHisPheAspGlnIleSerValSer 294
QY 871 TCT-----GTGCTGGAGTGACTGACACACAGGGCAA-----GAGAACAGT 912
Db 295 SerGlnTyrSerMetAsnTrpSerAlaGluArgSerArgLeuAsnTyrValGluAsnGly 314
QY 913 TGGAAACCCAAAAGCCAGGCTGAAAGAACCTGGACCGCTTGGGCTGCTTTTGGCACT 972
Db 315 TrpThrProGlyGluAspThrValLys----- 323
QY 973 CATGAATACCACTGGTTCACAAATAGATTGTAAGGAAAGAAAGAAATAACAGGCATTATA 1032
Db 324 -----GluTrpIleGlnValAspLeuGluAsnLeuArgPheValSerGlyIleGly 340
QY 1033 ACCACTGGA-----TCCACCATGGTGAGCACAATTAATGCTGCTGCTACAGATC 1086
Db 341 ThrGlnGlyAlaIleSerLysGluThrLysLysLysTyrPheValLysSerTyrLysVal 360
QY 1087 CTGTACAGTGTATGGGCAGAAATGGACTGTGTACAGAGAGCCGTGGTGAGCAAGAT 1146
Db 361 AspileSerSerAsnGlyGluAspTrpIleThrLeuLysAspGly-----AsnLysHis 378
QY 1147 AGATATTTTCAAGAAACAAGATTTATCACCAGGATGTGCTAATACTTTTGGCCACCA 1206
Db 379 LeuValPheThrGlyAsnThrAspAlaThrAspValTyrArgProPheSerLysPro 398
QY 1207 ATTATTGACGCTTTATTATAGTGAATCTACCAATGGCAGCAGAAATGGCATGAAA 1266
Db 399 ValIleThrArgPheValArgLeuArgProValThrTrpGluAsnGlyIleSerLeuArg 418
```

```
QY 1267 ATGAGACTGCTCGGATGTCAGTTTATTCCTAAAGGTGCTCTCCAAACTTACTACACCT 1326
Db 419 PheGluLeuTyrGlyCys-----LysIleThrAspTyr 429
QY 1327 CCACCTCTCTCGGAACAGCAATGACCTC-----AAAACACTACAGCC 1368
Db 430 ProCysSerArgMetLeuGlyMetValSerGlyLeuLeuSerAspSerGlnIleThrAla 449
QY 1369 CTTCCAAAATAGCCAAAGTCGTGCCCAAAATTTACGCAACACTACACCTCGCAGT 1428
Db 450 SerSerGlnValAspArgAsnTrpValProGluLeuAlaArgLeuValThrSerArgSer 469
QY 1429 -----AGCAATGAATTTCTCCACAGACAGAA 1455
Db 470 GlyTrpAlaLeuProProSerAsnThrHisProTyrThrLysGlu 484

RESULT 9
FAS_PIG STANDARD; PRT; 2258 AA.
ID FA5_PIG
AC O9GLP1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C
RP DOMAINS.
RC TISSUE=Liver;
RX MEDLINE=21121490; PubMed=11229814;
RA Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,
RA Kim H.K.W.;
RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional
RT protein modeling of membrane binding sites and comparative anatomy of
RT domains."
RL Cell. Mol. Life Sci. 58:148-159(2001).
CC -I- FUNCTION: Coagulation factor V is a cofactor that participates
CC with factor Xa to activate prothrombin to thrombin.
CC -I- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC chain, noncovalently bound. The interaction between the two chains
CC is calcium-dependent.
CC -I- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1
CC and C2 may be involved in membrane binding.
CC -I- PTM: Thrombin activates factor V proteolytically to the active
CC cofactor, factor Va (formation of a heavy chain at the N-
CC terminus and a light chain at the C-terminus).
CC -I- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -I- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -I- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL: AF191308; AAC28381.1;
CC HSP: P12259; IC2F.
CC InterPro: IPR001117; Cu-oxidase.
CC InterPro: IPR000421; FA58_C.
CC Pfam: PF00394; Cu-oxidase; 3.
CC Pfam: PF00754; F5_F8_type_C; 2.
CC SMART: SM00231; FA58C; 2.
CC PROSITE: PS00079; MULTICOPPER_OXIDASE1; 2.
CC PROSITE: PS01285; FA58C_1; 2.
```

DR PROSITE; PS01286; FA58C_2; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat. 1 22 POTENTIAL.
FT SIGNAL 23 2258 COAGULATION FACTOR V.
FT CHAIN 23 737 HEAVY CHAIN (BY SIMILARITY).
FT CHAIN 23 737 ACTIVATION PEPTIDE (CONNECTING REGION)
FT PEPTIDE 738 1611 (BY SIMILARITY).
FT CHAIN 1612 2258 LIGHT CHAIN (BY SIMILARITY).
FT DOMAIN 30 329 F5/8 TYPE A 1.
FT DOMAIN 30 329 PLASTOCYANIN-LIKE 1.
FT DOMAIN 203 329 PLASTOCYANIN-LIKE 2.
FT DOMAIN 348 683 F5/8 TYPE A 2.
FT DOMAIN 348 525 PLASTOCYANIN-LIKE 3.
FT DOMAIN 535 683 PLASTOCYANIN-LIKE 4.
FT DOMAIN 691 1611 B.
FT DOMAIN 1168 1539 41 X 9 AA APPROXIMATE TANDEM REPEATS OF
T-L-S-P-D-L-[GS]-[HQ]-T.
FT REPEAT 1168 1176 1.
FT REPEAT 1177 1185 2.
FT REPEAT 1186 1194 3.
FT REPEAT 1195 1203 4.
FT REPEAT 1204 1212 5.
FT REPEAT 1213 1221 6.
FT REPEAT 1222 1230 7.
FT REPEAT 1231 1239 8.
FT REPEAT 1240 1248 9.
FT REPEAT 1249 1257 10.
FT REPEAT 1258 1266 11.
FT REPEAT 1267 1275 12.
FT REPEAT 1276 1284 13.
FT REPEAT 1285 1293 14.
FT REPEAT 1294 1302 15.
FT REPEAT 1303 1311 16.
FT REPEAT 1312 1320 17.
FT REPEAT 1321 1329 18.
FT REPEAT 1330 1338 19.
FT REPEAT 1339 1347 20.
FT REPEAT 1348 1356 21.
FT REPEAT 1357 1365 22.
FT REPEAT 1366 1374 23.
FT REPEAT 1375 1383 24.
FT REPEAT 1384 1392 25.
FT REPEAT 1393 1401 26.
FT REPEAT 1402 1410 27.
FT REPEAT 1411 1419 28.
FT REPEAT 1420 1428 29.
FT REPEAT 1429 1437 30.
FT REPEAT 1438 1446 31.
FT REPEAT 1447 1455 32.
FT REPEAT 1456 1464 33.
FT REPEAT 1465 1473 34.
FT REPEAT 1474 1482 35.
FT REPEAT 1483 1491 36.
FT REPEAT 1492 1500 37.
FT REPEAT 1501 1509 38.
FT REPEAT 1510 1518 39.
FT REPEAT 1519 1527 40.
FT REPEAT 1531 1539 41.
FT DOMAIN 1616 1785 F5/8 TYPE A 3.
FT DOMAIN 1616 1785 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1795 1941 PLASTOCYANIN-LIKE 6.
FT DOMAIN 1942 2095 F5/8 TYPE C 1.
FT DOMAIN 2100 2355 F5/8 TYPE C 2.
FT SITE 737 738 CLEAVAGE (BY THROMBIN) (BY
SIMILARITY).
FT SITE 1029 1030 CLEAVAGE (BY THROMBIN) (BY
SIMILARITY).
FT SITE 1611 1612 CLEAVAGE (BY THROMBIN) (BY
SIMILARITY).
FT DISULFID 167 193 PROBABLE.
FT DISULFID 499 525 PROBABLE.
FT DISULFID 1759 1785 PROBABLE.

FT DISULFID 1941 2095 BY SIMILARITY.
FT DISULFID 2100 2255 BY SIMILARITY.
FT MOD_RES 692 SULFATION (POTENTIAL).
FT MOD_RES 696 SULFATION (POTENTIAL).
FT MOD_RES 724 SULFATION (POTENTIAL).
FT MOD_RES 726 SULFATION (POTENTIAL).
FT MOD_RES 745 SULFATION (POTENTIAL).
FT MOD_RES 1560 SULFATION (POTENTIAL).
FT MOD_RES 1576 SULFATION (POTENTIAL).
FT MOD_RES 1581 SULFATION (POTENTIAL).
FT MOD_RES 1584 SULFATION (POTENTIAL).
FT MOD_RES 1588 SULFATION (POTENTIAL).
FT MOD_RES 1631 SULFATION (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .)
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .)
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .)
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .)
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .)
FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .)
FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .)
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .)
FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .)
FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .)
FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .)
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .)
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .)
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .)
FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1048 1048 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1057 1057 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1066 1066 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1480 1480 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1537 1537 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1597 1597 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .)
FT CARBOHYD 1886 1886 N-LINKED (GLCNAC. . .)
FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. . .)
FT CARBOHYD 2243 2243 N-LINKED (GLCNAC. . .)
SQ SEQUENCE 2258 AA; 256078 MW; 9159B9E0076A2ACC CRC64;

Alignment Scores:

Pred. No.: 5,92e-18 Length: 2258
Score: 362.00 Matches: 79
Percent Similarity: 58.43% Conservative: 18
Best Local Similarity: 47.59% Mismatches: 59
Query Match: 8.74% Indels: 10
DB: 1 Gaps: 3

US-10-060-830-1 (1-2280) x FA5_PIG (1-2258)

QY 799 AAGACAAGTGGATGTATGGAACACTGGGGATGGAGTGTGTGTATCCGGATCTCAA 858
Db 2096 GluValAsnGlyCysSerThrProLeuGlyMetGluSerGlyAsnIleLysAsnGluIn 2115
QY 859 ATAACAGCATCATCTGCTGGAG-----TGGACTGACCACACAGGGCAAGAACACT 912
Db 2116 IleThrAlaSerSerPheLysSerTrpTrpGlyAspTr----- 2129
QY 913 TGAACACCCAAAAAGCCAGCTGAAAAACCTGGA-----CCGCTTGGGCTGCTTTT 966
Db 2130 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2149
QY 967 GCCACTGTGAATACCAAGTGGTTACAAATAGATTGAATTAAGAAAGAAATAACAGGC 1026
Db 2150 AlaAsnAsnAsnGlnTrpLeuGlnIleAspLeuLeuLysIleLysIleThrAla 2169
QY 1027 ATATAACCACTGGATCCACCATGGTGGAGCACAAATATTATGTCTGCTACAGAATC 1086
Db 2170 IleThrThrGlnGlyCysLysSerLeuSerSerGluMetTrpValLysArgTrpThrIle 2189

QY 1087 CTGTACAGTGTATGGCAGAAATGGACTGTGTACAGAGAGCCTGGTGTGGAGCAAGAT 1146
 DB 2190 GlnTyrSerAspArgGlyValGluIrrPlysserTyrArgGluLysSerSerMetValAsp 2209
 QY 1147 AAGATATTTCAGGAAACAAAGATTATCACCAGGATGTCGCTAATAACTTTTGGCCACCA 1206
 DB 2210 LysIlePheGluGlyAsnAsnAsnIleLysGlyHisValLysAsnPhePheAsnProPro 2229
 QY 1207 ATTATTGACGTTTATTAGTAGTACCTACCCAGTACCCAGGAGCAGAGAAATTCGCAAGAA 1266
 DB 2230 IleIleSerArgPheIleArgIleIleProLysMetTrpAsnGlnSerIleAlaLeuArg 2249
 QY 1267 ATGGAGCTGCTCGGATGT 1284
 DB 2250 LeuGluLeuPheGlyCys 2255

RESULT 10
 FA5_BOVIN
 ID FA5_BOVIN STANDARD; PRT; 2211 AA.
 AC Q28107; Q28108;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Coagulation factor V precursor (Activated protein C cofactor).
 GS F5.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=92147638; PubMed=1737753;
 RA Gualto E.R., Esmo C.T., Mann K.G., Macgillivray R.T.;
 RT "The complete cDNA sequence of bovine coagulation factor V.";
 RL J. Biol. Chem. 267:2971-2978(1992).
 CC 1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC 1- SUBUNIT: Factor Va is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC 1- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14
 CC AA repeats.
 CC 1- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC 1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
 CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY (BY SIMILARITY).
 CC 1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
 CC 2 PLASTOCYANIN-LIKE REPEATS.
 CC 1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC 1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; M81440; AAA30512.1; -;
 EMBL; M81441; AAA30513.1; -;
 DR HSPF; P12259; ICZT.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR000421; FA58.C.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
 DR PROSITE; PS01285; FA58C.1; 2.
 DR PROSITE; PS01286; FA58C.2; 2.

KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
 KW Repeat.
 FT SIGNAL. 1 28
 FT CHAIN 29 2211
 FT CHAIN 29 741
 FT PEPTIDE 742 1564
 FT CHAIN 1565 2211
 FT DOMAIN 30 327
 FT DOMAIN 30 193
 FT DOMAIN 203 327
 FT DOMAIN 348 686
 FT DOMAIN 348 525
 FT DOMAIN 535 686
 FT DOMAIN 696 1564
 FT SIMILAR 899 915
 FT DOMAIN 1124 1151
 FT REPEAT 1124 1137
 FT REPEAT 1138 1151
 FT DOMAIN 1188 1453
 FT REPEAT 1188 1196
 FT REPEAT 1197 1205
 FT REPEAT 1206 1214
 FT REPEAT 1215 1223
 FT REPEAT 1224 1232
 FT REPEAT 1233 1241
 FT REPEAT 1242 1250
 FT REPEAT 1251 1259
 FT REPEAT 1260 1268
 FT REPEAT 1269 1277
 FT REPEAT 1278 1286
 FT REPEAT 1287 1295
 FT REPEAT 1296 1304
 FT REPEAT 1305 1313
 FT REPEAT 1314 1322
 FT REPEAT 1323 1331
 FT REPEAT 1332 1340
 FT REPEAT 1341 1349
 FT REPEAT 1350 1358
 FT REPEAT 1359 1367
 FT REPEAT 1368 1376
 FT REPEAT 1377 1385
 FT REPEAT 1386 1394
 FT REPEAT 1395 1403
 FT REPEAT 1404 1412
 FT REPEAT 1413 1421
 FT REPEAT 1422 1430
 FT REPEAT 1431 1439
 FT REPEAT 1440 1444
 FT REPEAT 1445 1453
 FT REPEAT 1454 1462
 FT DOMAIN 1569 1738
 FT DOMAIN 1748 1890
 FT DOMAIN 1894 2048
 FT DOMAIN 2053 2208
 FT SITE 741 742
 FT SITE 1034 1035
 FT SITE 1564 1565
 FT DISULFID 167 193
 FT DISULFID 499 525
 FT DISULFID 1712 1738
 FT DISULFID 1894 2048
 FT DISULFID 2053 2208
 FT MOD_RES 697 697
 FT MOD_RES 701 701
 FT MOD_RES 730 730
 FT MOD_RES 1513 1513
 FT MOD_RES 1529 1529
 FT MOD_RES 1537 1537
 FT MOD_RES 1541 1541
 FT CARBOHYD 225 225
 FT CARBOHYD 239 239

POTENTIAL.
 COAGULATION FACTOR V.
 HEAVY CHAIN (BY SIMILARITY).
 ACTIVATION PEPTIDE (CONNECTING REGION)
 (BY SIMILARITY).
 LIGHT CHAIN (BY SIMILARITY).
 F5/8 TYPE A 1.
 PLASTOCYANIN-LIKE 1.
 PLASTOCYANIN-LIKE 2.
 F5/8 TYPE A 2.
 PLASTOCYANIN-LIKE 3.
 PLASTOCYANIN-LIKE 4.
 B.
 TO 17 AA REPEATS IN HUMAN FA5.
 2 X 14 AA TANDEM REPEATS.
 1-1.
 1-2.
 30 X 9 AA APPROXIMATE TANDEM REPEATS OF
 [AS]-L-S-P-D-[LP]-(GS)-Q-(TE).
 2-1.
 2-2.
 2-3.
 2-4.
 2-5.
 2-6.
 2-7.
 2-8.
 2-9.
 2-10.
 2-11.
 2-12.
 2-13.
 2-14.
 2-15.
 2-16.
 2-17.
 2-18.
 2-19.
 2-20.
 2-21.
 2-22.
 2-23.
 2-24.
 2-25.
 2-26.
 2-27.
 2-28.
 2-29 (PARTIAL).
 2-30.
 F5/8 TYPE A 3.
 PLASTOCYANIN-LIKE 5.
 PLASTOCYANIN-LIKE 6.
 F5/8 TYPE C 1.
 F5/8 TYPE C 2.
 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 PROBABLE.
 PROBABLE.
 BY SIMILARITY.
 BY SIMILARITY.
 SULFATION (POTENTIAL).
 SULFATION (POTENTIAL).
 SULFATION (POTENTIAL).
 SULFATION (POTENTIAL).
 SULFATION (POTENTIAL).
 SULFATION (POTENTIAL).
 SULFATION (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	297	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	382	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	460	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	553	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	587	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	745	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	756	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	774	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	780	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	902	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	952	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	964	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1044	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1053	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1062	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1071	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1078	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1094	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1451	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1490	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1550	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1690	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1839	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1997	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2196	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	VARIANT	587	NETLPA -> T (IN VARIANT 2).	
SQ	SEQUENCE	2211 AA; 248981 MW; CBBF90B738667C45 CRC64;		
Alignment Scores:				
Pred. No.:	6,94e-18	Length:	2211	
Score:	361.00	Matches:	79	
Percent Similarity:	57.83%	Conservative:	17	
Best Local Similarity:	47.59%	Mismatches:	60	
Query Match:	8.71%	Indels:	10	
DB:	1	Gaps:	3	
US-10-060-830-1 (1-2280) x FA5_BOVIN (1-2211)				
Qy	799	AAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGATCGCGGACCTCA	858	
Db	2049	GluValAsnGlyCysSerThrProLeuGlyMetGluSerGlyLysIleGluAsnLysGln	2068	
Qy	859	ATACAGCATCATGCTGCTGGAG-----TGGACTGACCCACACAGGGCAAGAGAAGT	912	
Db	2069	IleThrAlaSerSerPheLysLysSerTrpTrp-----GlyAsnTyr	2082	
Qy	913	TGGAACCCAAAAGCCAGCTGAAACCTGGA-----CCGCTTGGGCTGCTTT	966	
Db	2083	TrpGluProPheLeuAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys	2102	
Qy	967	GCCACTGATGAATACCATGCTTACAAATAGATTGTAATAGAGAAAGAAATAACAGGC	1026	
Db	2103	AlaAsnAsnAsnAsnGlnTrpLeuGlnIleAspLeuLeuLysIleLysIleThrAla	2122	
Qy	1027	ATTATAACCATGGATCCACCATCGGAGCACAATTACTATGCTCTGCTCCACAGATC	1086	
Db	2123	IleValThrGlnGlyCysLysSerLysSerGluMetTyrValLysSerTyrThrIle	2142	
Qy	1087	CTGTACAGTGTATGGCGAAGATGACTGTGTACAGAGCCCTGGTGTGGACCAAT	1146	
Db	2143	HisTyrSerAspGlnGlyThrAspTrpLysProTyrArgGluLysSerSerMetValAsp	2162	
Qy	1147	AAGATATTTCAGGAACAAGATATACACAGGATGTCGTAATACTTTTGGCACA	1206	
Db	2163	LysIlePheGluGlyAsnAsnValArgGlyHisValLysAsnPheAsnProPro	2182	
Qy	1207	ATTATTGCACGCTTTATTAGAGTGAATCCTACCAATGGCGAGCAAAATGCCATGAAA	1266	
Db	2183	IleIleSerArgPheIleArgIleIleProLysThrTrpAsnGlnSerIleAlaLeuArg	2202	
Qy	1267	ATGAGCTGCTCGGATG	1284	
Db	2203	LeuGluLeuPheGlyCys	2208	

RESULT 11	
MFQM_BOVIN	
AC	Q95114; Q27959; P79344; STANDARD; PRT; 427 AA.
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DE	Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE	(MGP57/53) (PAS-6/PAS-7 glycoprotein) (MFQM) (Sperm surface protein SP47) (BP47) (Components 15/16).
GN	MFG-E8
OS	Bos taurus (Bovine).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC	Bovidae; Bovinae; Bos.
OX	NCBI_TaxID=9913;
RN	[1]
RC	SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC	STRAIN-Holstein; TISSUE-Mammary gland;
RX	MEDLINE=97008954; PubMed=8856064;
RA	Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA	Petersen T.E.;
RT	"Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat globules.";
RT	Eur. J. Biochem. 240:628-636(1996).
RN	[2]
RP	SEQUENCE OF 18-427 FROM N.A.
RC	TISSUE-Mammary gland;
RX	MEDLINE=96125736; PubMed=8541316;
RA	Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA	Matsuda T.;
RT	"Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal antibodies raised against bovine milk fat globule membrane.";
RT	Biochim. Biophys. Acta 1245:385-391(1995).
RN	[3]
RP	SEQUENCE OF 19-427 FROM N.A.
RC	TISSUE-Testis;
RA	Ensslin M.A.;
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN	[4]
RP	SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC	TISSUE-Milk;
RX	MEDLINE=93250576; PubMed=8485470;
RA	Mather I.H., Banghart L.R., Lane W.S.;
RT	"The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein containing epidermal growth factor-like and factor V/VIII-like sequences.";
RT	Biochem. Mol. Biol. Int. 29:545-554(1993).
CC	-1- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-BINDING PROTEIN.
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN.
CC	-1- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
CC	-1- PTM: THE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC	-1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X91895; CA62997.1; -.
DR	EMBL; S80643; AAB35894.2; -.

AC P12259; Q14285;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=92232668; PubMed=1567832;
RX Cripe L.D., Moore K.D., Kane W.H.;
RA "Structure of the gene for human coagulation factor V.";
RL Biochemistry 31:3777-3785(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87260886; PubMed=3110773;
RA Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
Hewick R.M., Kaufman R.J., Mann K.G.;
RL "Complete cDNA and derived amino acid sequence of human factor V.";
Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
RN [3]
RP SEQUENCE OF 1-1600 FROM N.A.
RX MEDLINE=88107560; PubMed=2827731;
RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
RL "Cloning of cDNAs coding for the heavy chain region and connecting
region of human factor V, a blood coagulation factor with four types
of internal repeats.";
Biochemistry 26:6506-6514(1987).
RN [4]
RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
RX MEDLINE=86313665; PubMed=3092220;
RA Kane W.H., Davie E.W.;
RL "Cloning of a cDNA coding for human factor V, a blood coagulation
factor homologous to factor VIII and ceruloplasmin.";
Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RC TISSUE=Fibroblast;
RX MEDLINE=93203619; PubMed=8454869;
RA Shen N.L.L., Fan S.-T., Pyatt J., Graff R., Lapolla R.J.,
Edgington T.S.;
RL "The serine protease cofactor factor V is synthesized by
lymphocytes.";
J. Immunol. 150:2992-3001(1993).
RN [6]
RP SULFATION.
RX MEDLINE=94264012; PubMed=8204629;
RA Pittman D.D., Tomkinson K.N., Michnick D., Sellighsohn U.,
Kaufman R.J.;
RL "Posttranslational sulfation of factor V is required for efficient
thrombin cleavage and activation and for full procoagulant activity.";
Biochemistry 33:6952-6959(1994).
RN [7]
RP SULFATION.
RX MEDLINE=90366699; PubMed=2168225;
RA Hortin G.L.;
RL "Sulfation of tyrosine residues in coagulation factor V.";
Blood 76:946-952(1990).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
RX MEDLINE=20052169; PubMed=10586886;
RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
Ortel T.L., Bourneken G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
Fuentes-Prior P.;
RL "Crystal structures of the membrane-binding C2 domain of human
coagulation factor V.";
Nature 402:434-439(1999).
RN [9]
RP VARIANT APCR GLN-534.
RX MEDLINE=94217810; PubMed=8164741;
RA Bertina R.M., Koelman B.P.C., Koster T., Rosendaal F.R.,
Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
RL "Mutation in blood coagulation factor V associated with resistance to
activated protein C.";
Nature 369:64-67(1994).
RN [10]
RP VARIANT APCR GLY-334, AND VARIANT LYS-513.
RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
RL "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese.";
Blood 91:1135-1139(1998).
RN [11]
RP VARIANT APCR THR-334.
RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
RL "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
resistance to activated protein C.";
Blood 91:1140-1144(1998).
CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
chain, noncovalently bound. The interaction between the two chains
is calcium-dependent.
CC -1- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
REPEATS.
CC -1- PTM: Thrombin activates factor V proteolytically to the active
cofactor, factor Va (formation of a heavy chain at the N-
terminus and a light chain at the C-terminus).
CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
CC -1- DISEASE: OWEN PARAHEMOPHILIA, AN HEMORRHAGIC DIATHESIS, IS DUE
TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APCR
MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
IMPLANTATION.
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
2 PLASTOCYANIN-LIKE REPEATS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; L32779; AAB59401.1;
DR EMBL; L32755; AAB59401.1; JOINED.
DR EMBL; L32756; AAB59401.1; JOINED.
DR EMBL; L32757; AAB59401.1; JOINED.
DR EMBL; L32758; AAB59401.1; JOINED.
DR EMBL; L32759; AAB59401.1; JOINED.
DR EMBL; L32760; AAB59401.1; JOINED.
DR EMBL; L32761; AAB59401.1; JOINED.
DR EMBL; L32762; AAB59401.1; JOINED.
DR EMBL; L32763; AAB59401.1; JOINED.
DR EMBL; L32764; AAB59401.1; JOINED.
DR EMBL; L32765; AAB59401.1; JOINED.
DR EMBL; L32766; AAB59401.1; JOINED.
DR EMBL; L32767; AAB59401.1; JOINED.
DR EMBL; L32768; AAB59401.1; JOINED.
DR EMBL; L32769; AAB59401.1; JOINED.
DR EMBL; L32770; AAB59401.1; JOINED.
DR EMBL; L32771; AAB59401.1; JOINED.
DR EMBL; L32772; AAB59401.1; JOINED.
DR EMBL; L32773; AAB59401.1; JOINED.
DR EMBL; L32774; AAB59401.1; JOINED.
DR EMBL; L32775; AAB59401.1; JOINED.
DR EMBL; L32776; AAB59401.1; JOINED.
DR EMBL; L32777; AAB59401.1; JOINED.
DR EMBL; L32778; AAB59401.1; JOINED.
DR EMBL; M16967; AAA52424.1;

```

DR EMBL; M14335; AAB59532.1; -.
DR PIR; A25897; A25897.
DR PIR; A28028; A28028.
DR PDB; 1CZS; 26-NOV-99.
DR PDB; 1CZT; 26-NOV-99.
DR PDB; 1CZV; 26-NOV-99.
DR Genew; HGNC:3542; F5.
DR MIM; 227400; -.
DR MIM; 134400; -.
DR MIM; 188055; -.
DR MIM; 227310; -.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_P8_type_C; 2.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat; Polymorphism; Disease mutation; Thrombophilia; 3D-structure.
FT SIGNAL 1 28
FT CHAIN 29 2224
FT CHAIN 29 737
FT PEPTIDE 738 1573
FT CHAIN 1574 2224
FT CHAIN 30 329
FT DOMAIN 30 193
FT DOMAIN 203 329
FT DOMAIN 348 684
FT DOMAIN 348 526
FT DOMAIN 536 684
FT DOMAIN 692 1573
FT DOMAIN 895 928
FT REPEAT 895 911
FT REPEAT 912 928
FT REPEAT 1135 1148
FT SIMILAR 1135 1148
FT DOMAIN 1185 1501
FT REPEAT 1185 1193
FT REPEAT 1194 1202
FT REPEAT 1203 1211
FT REPEAT 1212 1220
FT REPEAT 1221 1229
FT REPEAT 1230 1238
FT REPEAT 1239 1247
FT REPEAT 1248 1256
FT REPEAT 1257 1265
FT REPEAT 1266 1274
FT REPEAT 1275 1283
FT REPEAT 1284 1292
FT REPEAT 1293 1301
FT REPEAT 1302 1310

Alignment Scores:
Pred. No.: 6,83e-17
Score: 347.00
Percent Similarity: 58.43%
Best Local Similarity: 45.78%
Query Match: 8.38%
DB: 1
Gaps: 3

US-10-060-830-1 (1-2280) x FA5_HUMAN (1-2224)
QY 799 AAGACAAGTGGATGATGGAACACTGGGATGGAGTCTGGTGTGATCGCGGATCCTCAA 858
Db 2062 GluValAsnGlyCysSerThrProLeuGlyMetGluAsnGlyLysIleGluAsnLysGln 2081
QY 859 ATAACAGCATCATCTGCTGCTGGAG-----TGGACTGACACACAGGCGCAAGAGAACAGT 912
Db 2082 IleThrAlaSerSerPheLysLysSerTrpTrpGlyAspTyr----- 2095
QY 913 TGGAAACCAAAAGCCAGGCTGAAAAACCTGGA-----CCGCTTGGCTGCTTTT 966

Db 2096 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2115
QY 967 GCCACTGATGATACCACTGGTTACAAATAGATTGAAATAGAGAAAGAAATAACAGGC 1026
Db 2116 AlaAsnAsnLysGlnTrpLeuGlnIleAspLeuLeuLysLysIleThrAla 2135
QY 1027 ATTATACCACTGGATCCACCATGGTGGAGCACAATTACTATGTCTGCTCCCTACAGAATC 1086
Db 2136 IleIleThrGlnGlyCysLysSerLeuSerSerGluMetTyrValLysSerTyrThrIle 2155
QY 1087 CTGTACAGTGTATGGGAGCAAAATGGACTGTGTACAGAGACCTGGTGTGGGCAAGAT 1146
Db 2156 HisTyrSerGluGlnGlyValGluTrpLysProTyrArgLeuLysSerSerMetValAsp 2175
QY 1147 AAGATATTCAAGGAAACAAGATTATCACAGGATGTCCTGAATAACTTTTGGCCACCA 1206
Db 2176 LysIlePheGluGlyAsnThrAsnThrLysGlyHisValLysAsnPhePheAsnProPro 2195
QY 1207 ATTATTGACACTTTTATTAGAGTGAATCCTACCAATGGCAGCAGAAATTCGCCATGAA 1266
Db 2196 IleIleSerArgPheIleArgValIleProLysThrTrpAsnGlnSerIleThrLeuArg 2215
QY 1267 ATGGAGCTGCTCGGATG 1284
Db 2216 LeuGluLeuPheGlyCys 2221

RESULT 13
MFGM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFG-E8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE-Breast, and Breast carcinoma;
RX MEDLINE=96213908; PubMed=8639264;
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
"Cloning and sequence analysis of human breast epithelial antigen
BA46 reveals an RGD cell adhesion sequence presented on an epidermal
growth factor-like domain.";
RT DNA Cell Biol. 15:281-286(1996).
[2]
SEQUENCE OF 170-387 FROM N.A.
TISSUE-Mammary gland;
RX MEDLINE=91371351; PubMed=1909932;
RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
Ceriani R.L.;
"A Mr 46,000 human milk fat globule protein that is highly expressed
in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
[3]
PARTIAL SEQUENCE, AND CHARACTERIZATION.
TISSUE-Milk;
RX MEDLINE=98194924; PubMed=9535276;
RA Guiffrida M.G., Cavalletto M., Giunta C., Conti A.,
Godovac-Zimmermann J.;
"Isolation and characterization of full and truncated forms of human
breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
[4]
SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=95342076; PubMed=10411933;
RA Haeggqvist B., Naeslund J., Sletten K., Westermarck G.T., Mucchiano G.,
Tjernberg L.O., Nordstedt C., Engstrom U., Westermarck P.;

```

RT "Medin: an integral fragment of aortic smooth muscle cell-produced
 RL lactadherin forms the most common human amyloid";
 RT Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE:97405885; PubMed:9260929;
 RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
 RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
 RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
 RT (RGD)-dependent cell adhesion";
 RL DNA Cell Biol. 16:861-869(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 CC SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
 CC -1- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
 CC MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
 CC -1- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
 CC AMINO ACID 264 AND 273.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U58516; AAC50549.1; -;
 DR EMBL; S56151; AAB19771.1; -;
 DR HSP; P08709; Ibf9.
 DR Genew; HGNC:7036; MFG8.
 DR MIM; 602281;
 DR SMART; SM00181; EGF_1;
 DR SMART; SM00231; FAS8C; 2.
 DR PROSITE; PS00022; EGF_1;
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01285; FAS8C_1; 2.
 DR PROSITE; PS01286; FAS8C_2; 2.
 KW Signal; glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
 FT SIGNAL 1 23
 FT CHAIN 24 387
 FT CHAIN 202 387
 FT CHAIN 268 317
 FT DOMAIN 24 67
 FT DOMAIN 70 225
 FT DOMAIN 230 387
 FT SITE 46 48
 FT SITE 27 38
 FT DISULFID 32 55
 FT DISULFID 57 66
 FT DISULFID 70 225
 FT DISULFID 212 216
 FT DISULFID 230 387
 FT CARBOHYD 238 238
 FT CARBOHYD 325 325
 FT CARBOHYD 329 329
 FT CARBOHYD 350 350
 SQ SEQUENCE 387 AA; 43123 MW; 2EE6571DEC83782D CRC64;
 Alignment Scores:
 Pred. No.: 9, 26e-17
 Score: 343.00
 Percent Similarity: 57.23%
 Best Local Similarity: 41.04%
 Query Match: 8.28%
 DB: 1
 Gaps: 3

US-10-060-830-1 (1-2280) x MFGM_HUMAN (1-387)
 QY 775 TTATCTACAAGTCTTTTACATTAAACACAAGTGGATGTTATGGAACACACCTGGGATGGAG 834
 Db 218 LtuArgPheGluLeuLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 237
 QY 835 TCTGGTGTGATCGCGGATCTCTCAAAATACACAGCATCATCTGTCTGGAG---TGGACTGAC 891
 Db 238 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerTyrLysThrTTPGlyLeu 257
 QY 892 CACACAGGGCAAGAAACAGTGTGAACACCAACCAAGCCAGGCTGAAAAACCTGGA--- 948
 Db 258 HisLeu-----PheSerTTPAsnProSerTyrAlaArgLeuAspLysGlnGlyAsn 274
 QY 949 ---CCGCTTGGCGTCTTTTGGCCACTGATGATACCACTGGTTTACAAATAGATTTCGAAT 1005
 Db 275 PheAsnAlaTTPValAlaGlySerTyrGlyAsnAspGlnTTPLeuGlnValaPLeuGly 294
 QY 1006 AAGGAAAGAAATPACAGGCAATTAACCACTGGATCCACCATGGTGGAGCACAATATC 1065
 Db 295 SerSerLysGluValThrGlyIleIleThrGlnGlyAlaArgAsnPheGlySerValGln 314
 QY 1066 TATGTCTGCTACAAATCCCTGACAGTATGATGGCAGAAATGGACTGTGTACAGA 1125
 Db 315 PheValAlaSerTyrLysValAlaTyrSerAsnAspSerAlaAsnTTPThrGlyTyrGln 334
 QY 1126 GAGCTGTGTGGAGCAAGATAGATATTTCAAGGAAACAAAGATTATCATCCAGGATGTG 1185
 Db 335 AspProArgThrGlySerSerLysIlePheProGlyAsnTTPAspAsnHisSerHisLys 354
 QY 1186 CGTAATAACTTTTCCCAACCAATTTATTCAGCTTTTATAGTGAATCCTTACCAATGG 1245
 Db 355 LysAsnLeuPheGluThrProIleLeuAlaArgTyrValArgIleLeuProValAlaTTP 374
 QY 1246 CAGCAGAAATGGCATGAAATGGAGCTGCTCGGATGT 1284
 Db 375 HisAsnArgIleAlaLeuArgLeuGluLeuLeuGlyCys 387
 RESULT 14
 MFGM_PIG
 ID MFGM_PIG STANDARD; PRT; 409 AA.
 AC P79385;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Lactadherin (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM) (Sperm
 DE surface protein SP47) (PP47).
 GN MFG8.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID:9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Ensslin M.A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
 CC BINDING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
 CC SPERMATOZOAN.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```
CC ----- EMBL; Y11683; CAA72379.1; -.
DR HSP; P00740; IEDM.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000421; FA58_C.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00734; F5_F8_Type_C; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW Glycoprotein; Repeat; EGF-like domain.
FT DOMAIN 2 41 EGF-LIKE 1.
FT DOMAIN 44 88 EGF-LIKE 2.
FT DOMAIN 91 247 F5/8 TYPE C 1.
FT DOMAIN 252 409 F5/8 TYPE C 2.
FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 6 17 BY SIMILARITY.
FT DISULFID 11 29 BY SIMILARITY.
FT DISULFID 31 40 BY SIMILARITY.
FT DISULFID 91 247 BY SIMILARITY.
FT DISULFID 234 238 BY SIMILARITY.
FT DISULFID 252 409 BY SIMILARITY.
FT CARBOHYD 41 41 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 409 AA; 45725 MW; B0C07AF80029927A CRC64;

Alignment Scores:
Pred. No.: 1.53e-16 Length: 409
Score: 340.00 Matches: 76
Percent Similarity: 56.59% Conservative: 27
Best Local Similarity: 41.76% Mismatches: 68
Query Match: 8.21% Indels: 11
DB: 1 Gaps: 5

US-10-060-830-1 (1-2280) x MFGM_PIG (1-409)
QY 759 ATCTGT-----GGTGGACACTTATCTACAGTCCTTTTACATTTAAGACAAGTGGATG 812
DB 233 llecysHisArgGlyCysThrLeuArgPheGlu-LeuLeuGlyCysGluLeuSerGlyC 252
QY 813 TTATGGAACACTGGGATGAGTCTGTGTGATCGCGATCTCTCAATAACAGCATCATC 872
DB 252 salagluProLeuGlyLeuLysAspAsnThrIleProAsnLysGlnIleThrAlaSerSe 272
QY 873 TGTGCTGAG---TGGACTGACCACACAGG---CAAGAGACACAGTTGGAAACCCAAAA 926
DB 272 rPheTyrArgThrTrp-----GlyLeuSerAlaPheSerTrpTyrProPheTy 288
QY 927 AGCCAGGCTGAAAAACCTGGACCG-----CCTTGGCTGCTTTTGGCACTGATGAATA 980
DB 288 rAlaArgLeuAspAsnGlnGlyLysPheAsnAlaTrpThrAlaGlnSerAsnSerAlaSe 308
QY 981 CCAGTGGTTACAAATAGATTGTAAGAAAGAAATAACAGGCATTATTAACACTGG 1040
DB 308 rGluTrpLeuGlnIleAspLeuGlySerGlnArgArgValThrGlyIleIleThrGlnG 328
QY 1041 ATCCACCATGTTGGAGCACAAATTACTATGTCCTGCTCCCTACAGATCTCTGACATGATGA 1100
DB 328 yAlaArgAspPheGlyHisIleGlnTyrValAlaAlaTyrLysValAlaTyrSerAspAs 348
QY 1101 TGGCAGAAATGGACTGTGTACAGAGCCCTGTGTGGAGCAAGTAAAGATATTTCRAAG 1160
DB 348 pGlyValSerTrpThrGluTyrArgAspGlnGlyAlaLeuGluGlyLysIlePheProGl 368
QY 1161 AAACAAGATTATCACCAGGATGTCGTAATAACTTTTGGCCACCAATTATTGACGCTTT 1220
DB 368 yAsnLeuAspAsnSerHisLysLysAsnMetPheGluThrProPheLeuThrArgPh 388
QY 1221 TATTAGATGTAACTCCATCCCATGGCAGCAGAAATATGCCATGAATGAGAGCTGCTCGG 1280
DB 1221 TATTAGATGTAACTCCATCCCATGGCAGCAGAAATATGCCATGAATGAGAGCTGCTCGG 1280
```

```
Db 388 eValArgIleLeuProValAlaTrpHisAsnArgIleThrLeuArgValGluLeuLeuG 408
QY 1281 ATGT 1284
DB 408 yCys 409
RESULT 15
MFGM_RAT STANDARD; PRT; 427 AA.
AC P70490;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-
DE acetyl GD3 ganglioside synthase) (AGS) (MFGM).
GN MFG-E8 OR AGS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RX MEDLINE=96374422; PubMed=8780713;
RA Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
RT "Cloning and expression of cDNA for O-acetylation of GD3
RT ganglioside."
RL Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO
CC PARTICIPATE IN THE O-ACETYLTATION OF GD3 GANGLIOSIDE SIALIC ACID.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D84068; BAA12210.1; -.
DR HSP; P00740; IEDM.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR000421; FA58_C.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00734; F5_F8_Type_C; 2.
DR PRINTS; PR001010; EGFBL00D.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 427 LACTADHERIN.
FT DOMAIN 24 61 EGF-LIKE 1.
FT DOMAIN 64 108 EGF-LIKE 2.
FT DOMAIN 111 267 F5/8 TYPE C 1.
FT DOMAIN 272 427 F5/8 TYPE C 2.
FT DISULFID 28 39 BY SIMILARITY.
FT DISULFID 33 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 96 BY SIMILARITY.
FT DISULFID 98 107 BY SIMILARITY.
FT DISULFID 111 267 BY SIMILARITY.
FT DISULFID 254 256 BY SIMILARITY.
FT DISULFID 272 427 BY SIMILARITY.
```


DR EMBL: AF387547; AAL30178.1; --
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR000421; FAS8_C.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR PROSITE: PS01180; CUB; 1.
 KW DNA-binding; Transcription regulation.
 SQ SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;

Alignment Scores:
 Pred. No.: 0 Length: 775
 Score: 4017.00 Matches: 754
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 96.96% Indels: 0
 DB: 4 Gaps: 0

US-10-060-830-1 (1-2280) x Q96PD2 (1-775)

QY 1 GCCGCGCGCGCGCTGGCGCGCGCTCCCGCTCTCCGCTCCCTCCCTCCCTCCCAAC 60
 DB 22 AlaAlaAlaProAlaAlaProAlaAlaAlaProLeuSerArgSerLeuProCysSerAsn 41
 QY 61 TCCTCCTCCTTCTCCATGCCCTCTGCTCCTCCTGCTCTTACTTGTCTCCTGCTCCTGCTC 120
 DB 42 SerSerSerPheSerMetProLeuPheLeuLeuLeuLeuValLeuLeuLeuLeuLeu 61
 QY 121 GAGGACGCTGGAGCCGACGAGGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGT 180
 DB 62 GluAspAlaGlyAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 81
 QY 181 GGAACCCCTACATCCATAAATACCACAGACCTATCCACAGACACTATCCAAATGGAATGGAATGG 240
 DB 82 GlyThrLeuThrSerIleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTyrP 101
 QY 241 GAGATCCGTGTAAGATGGGAGAGAGTTCGCATCAAAATTTGGTGTGACTTTGACATTGAA 300
 DB 102 GluIleArgValLysMetGlyGluArgValArgIleLysPheGlyAspPheAspIleGlu 121
 QY 301 GATTCGTATCTGTACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 360
 DB 122 AspSerAspSerCysHisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArg 141
 QY 361 ACTGAAATAGGCAATACTGTGCTGGGTGGGTTGCAATGGAATGGAATGGAATGGAATGGAAT 420
 DB 142 ThrGluIleGlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLys 161
 QY 421 GGCAATGAATCACAATCTGCTTCAATGATGGATGCCATCCATGTTCTGGACGCGGATTTTGTG 480
 DB 162 GlyAsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeu 181
 QY 481 GCCTCATCTCTGTATAGATAAACAATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 540
 DB 182 AlaSerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsn 201
 QY 541 TTTTGGACCTGACTCAGTAAGTACGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
 DB 202 PheLeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGlu 221
 QY 601 ATATCTGGAACAATTCCTCATGATATAGATTCCTCGCATTTGTGATGCTGCTGCTGCTGCTG 660
 DB 222 IleSerGlyThrIleProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyVal 241
 QY 661 CATGACGAGTATGTCTCAACACGTTGGCGGCCCAATCAGTTGTTGTAATTAGTAAGGT 720
 DB 242 HisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValValIleSerLysGly 261
 QY 721 ATTCCCTATTATGAAGTCTTTGGCTAACACGCTCACTCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 DB 262 IleProTyrTyrGluSerSerLeuAlaAsnValThrSerValValGlyHisLeuSer 281
 QY 781 ACAAGCTCTTTTACATTTTAAGACAAGTGTGATGTTATGGAACACTGGGATGGAGTCTGGT 840

DB 282 ThrSerLeuPheThrPheLeuThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGly 301
 QY 841 GTGATCGCGATCCTCAATAACAGCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 DB 302 ValIleAlaAspProGlnIleThrAlaSerSerValLeuGluThrThrAspHisThrGly 321
 QY 901 CAAGAGAACACAGTTGGAAACCCAAAAAGCCAGGCTGAAAAAACCCTGGACCCCTGGCTG 960
 DB 322 GlnGluAsnSerTyrPheLysLysAlaArgLeuLysLysProGlyProProtrPala 341
 QY 961 GCTTTTGGCCACTGATGAATACCAAGTGTGTACAAATAGATTGTAATAAGGAAAAAATA 1020
 DB 342 AlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIle 361
 QY 1021 ACAGGCATTATACCACTGGATCCACCATGGTGGAGGACAAATTACTATGCTGCTGCTGCT 1080
 DB 362 ThrGlyIleIleThrThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyr 381
 QY 1081 AGAATCCTGTACAGTGTATGGCAGAAATGGAGTGTGTACAGAGAGCCCTGGTGTGGAG 1140
 DB 382 ArgIleLeuTyrSerAspAspGlyGlnLysTyrThrValTyrArgGluProGlyValGlu 401
 QY 1141 CAAGATAAGATATTTCAGGAACAAGAATATCACCAGGATGTCGCTAATAACTTTTGT 1200
 DB 402 GlnAspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeu 421
 QY 1201 CCACCAATTATGCACGTTTATTAGAGTGAATCCTCCCAATGGCAGCAGCAAAATGCC 1260
 DB 422 ProIleIleAlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAla 441
 QY 1261 ATGAAATGGAGTGTGCTCGGATGTCAAGTATTCTCAAGGCTGCTCCCAAACTTACT 1320
 DB 442 MetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThr 461
 QY 1321 CAACCTCCACCTCTCGGAACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAAATA 1380
 DB 462 GlnProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProLysIle 481
 QY 1381 GCCAAAGGTGCTGCCCAAAATTTACGCAACCACTACCACTCGCAGTAGCAATGAATTT 1440
 DB 482 AlaLysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPhe 501
 QY 1441 CCTCCACAGACAGAACAAACCTGCCAGTCTGATATCAGAAATACCTACCTCACTCCA 1500
 DB 502 ProAlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThrPro 521
 QY 1501 AATGTAACCAAGATGTAGCGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 1560
 DB 522 AsnValThrLysAspValAlaLeuAlaAlaValLeuValProValLeuValMetValLeu 541
 QY 1561 ACTACTCTCATTCATATTAGTGTGCTTGGCACTGGAGAACAGAGAAAGAAACT 1620
 DB 542 ThrThrLeuIleLeuIleLeuValCysAlaTrpHisTrpArgAsnArgLysLysThr 561
 QY 1621 GAAGGCACCTATGACTTACCTTACTGGACCGGGCAGGTTGGTGGAAAGGAATGAAGCAG 1680
 DB 562 GluGlyThrTyrAspLeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLysGln 581
 QY 1681 TTTCTTCTGCAAAAGCAGTGGACCATGAGGAAACCCAGTTCGCTATAGCAGCAGCAAA 1740
 DB 582 PheLeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSerSerSerGlu 601
 QY 1741 GTTAATCACCTGATGCCAAGAGAGTCCACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 DB 602 ValAsnHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyr 621
 QY 1801 GCTCAGCCACTGGTAGGAGGAATTCCTGTACACTTCATCAAGATCTACCTTTTAAACA 1860
 DB 622 AlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysPro 641
 QY 1861 CAAGAAGGAAAGACGAGGCTATCAGACCTTAGATCTCTACAACCTCACCAGGCGAGAA 1920

```

Db 642 GluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGlu 661
QY 1921 GTTATCATGCTATGCTGAACCACTCCCAATTACGGGGCTGAGTATGCAACCCCAATC 1980
Db 662 ValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIle 681
QY 1981 ATCATGACATGTCAGGCGACCCACAACTTCAGTTGGTCAGCCCTCCACATCCACTTTC 2040
Db 682 IleMetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPhe 701
QY 2041 AAGCTACGGGGAACCAACCTCCCCCACTAGTGGGAACCTTACAATACACTTCTCTCCAGG 2100
Db 702 LysAlaThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArg 721
QY 2101 ACTCACAGCTCTCTCTCAGCCAGCCAGCATGATACCCCGAAGCTGGGAAGCAGGT 2160
Db 722 ThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGly 741
QY 2161 CTACTGCCCGACGAAATGGTGTACCAAGTGCACAGACACACAAAGAGTATCAGGA 2220
Db 742 LeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGly 761
QY 2221 GCAGGAAGGATGGGGAATGTGATGTTTAAAGAAATCCCT 2262
Db 762 AlaGlyArgAspGlyGluCysAspValPheLysIleLeu 775

RESULT 2
Q8TDX2 PRELIMINARY; PRT; 775 AA.
AC Q8TDX2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CLCP1.
GN CLCP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Koshikawa K., Osada H., Kozaki K., Konishi H., Masuda A.,
RA Tamematsu Y., Mitsudomi T., Nakao A., Takahashi T.;
RT "Significant up-regulation of a novel gene, CLCP1, in a highly
RT metastatic lung cancer subline as well as in lung cancers in vivo.";
RL Oncogene 0:0-0(2002).
DR EMBL; AB073146; BAB91138.1; -
SQ SEQUENCE 775 AA; 85073 MW; 6D07223B21BE5A42 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 775
Score: 4013.00 Matches: 753
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.86% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-1 (1-2280) x Q8TDX2 (1-775)
QY 4 GCCGCCCCCGCTGGCGCGCTCCCTCTCCGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 63
Db 23 AlaAlaProAlaTyrAlaAlaLeuProLeuSerArgSerLeuProProCysSerAsnSer 42
QY 64 TCCCTCTCTCCATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 123
Db 43 SerSerPheSerMetProLeuPheLeuLeuLeuLeuValLeuLeuLeuLeuLeuLeu 62
QY 124 GACGCTGAGCCAGCAGGATGGATGGACACACTGACTAGGCGCTGAGAGTGA 183
Db 63 AspAlaGlyAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 82
QY 184 ACCCTTACATCCATAAATCAACCAAGACCTATCCCAACAGCACTGTTTGTGAATGGGAG 243

```

```

Db 83 ThrLeuThrSerIleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGlu 102
QY 244 ATCCGTGFAAAGATGGGAGAGAGAGTTCGCATCAAAATTTGGTGTACTTTGACATTGAAGAT 303
Db 103 IleArgValLysMetGlyGluArgValArgIleLysPheGlyAspPheAspIleGluAsp 122
QY 304 TCTGATCTCTGCTCACTTTTAAATTTACTTGAATTTATATGAATTTGGAGTGCAGCAGAACT 363
Db 123 SerAspSerCysHisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThr 142
QY 364 GAATAGGCAAAATCTGTGCTGGGTTGCAATGAACCAATTCATTAATGAATCAAAAGGC 423
Db 143 GluIleGlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGly 162
QY 424 AATGAATACATCTGCTGTCATGAGTGAATCCATGTTCTGGACGGGATTTTGGCC 483
Db 163 AsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAla 182
QY 484 TCATACTCTGTATAGATAAAACAAGATCTAATTAATTTTGGACACTGCATCCAATTTT 543
Db 183 SerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPhe 202
QY 544 TTGGAACCTGAGTTTCAGTAAGTACCTGCCAGCTGCTGCTGCTCTCTCTCTCTCTCTCT 603
Db 203 LeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIle 222
QY 604 TCTGGACAATTCCTCATGATATAGAGATTCCTCGCATTCATGCTGCTGCTGCTGCTGCT 663
Db 223 SerGlyThrIleProHisGlyTyrArgAspSerProLeuCysMetAlaGlyValHis 242
QY 664 GCAGGAGTAGTGTCAACACACGCTGGCGGCCAAATCACTGTTGTAATTAAGTAAGTATT 723
Db 243 AlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValIleSerLysGlyIle 262
QY 724 CCTATTATGAAGTTCTTTGGCTAACACGTCATCATCTGCTGGGGGACACTTATCTACA 783
Db 263 ProTyrTyrGluSerSerLeuAlaAsnValThrSerValValGlyHisLeuSerThr 282
QY 784 AGTCTTTTACATTTAAGACAAGTGGATGTTATGAACACACCTGGGATGGAGTCTGGTGTG 843
Db 283 SerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyVal 302
QY 844 ATCCGCGGATCTCAAAATCAACAGCATCATCTGCTGGAGTGGAGTGCACACACAGGCAA 903
Db 303 IleAlaAspProGlnIleThrAlaSerSerValLeuGluTyrThrAspHisThrGlyGln 322
QY 904 GAGAACAGTTGGAAACCCCAAGCCAGCTGAGTGAACCAACCTGGACCCCTGGGCTGCT 963
Db 323 GluAsnSerTyrLysProLysAlaArgLeuLysLysProGlyProProTrpAlaAla 342
QY 964 TTTGCCACTGATGAATACACAGTGGTTACAAATAGATTTGAATAAGGAAAGAAATAACA 1023
Db 343 PheAlaThrAspGluTyrGlnThrLeuGlnIleAspLeuAsnLysGluLysIleThr 362
QY 1024 GGCATTATPAACACTGGATCCACCATGCTGGAGCACAATTAATCTATGCTGCTGCTACAGA 1083
Db 363 GlyIleIleThrThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArg 382
QY 1084 ATCTGTACATGATGATGGCGAGAAATGGAGTGTGTACAGAGCCCTGGTGGAGCAA 1143
Db 383 IleLeuTyrSerAspAspGlyGlnLysThrThrValTyrArgGluProGlyValGluGln 402
QY 1144 GATAGATATTTCAGGAAACAAAGATTAATCACCAGGATGTCGTAATACTTTTGGCA 1203
Db 403 AspLysIlePheGlnGlnLysLysAspTyrHisGlnAspValArgAsnAspPheLeuPro 422
QY 1204 CCAATATTGACGCTTTTATTAGAGTGAATCCTACCCCAATGCGCAGCAAAATTTGCCATG 1263
Db 423 ProIleIleAlaArgPheIleArgValAsnProThrGlnTyrGlnGlnLysIleAlaMet 442
QY 1264 AAATGGAGCTGCTGGGATGTCAGTTTATCTTAAGCTGCTCTCTCTCTCTCTCTCTCTCT 1323

```


Db 175 ArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAsp 194
QY 529 ACTGCATCCAAATTTTGGAACTGAGTTCAGTAAGTACTGCCAGCTGGTGTCTCT 588
Db 195 ThrValSerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeu 214
QY 589 CCTTTTCTGAGATATCGGAACAATTCCTCATGGATATAGATTCTCTGCCATTGTC 648
Db 215 ProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerProLeuCys 234
QY 649 ATGGCTGGTGTGATGACAGAGTAGTGTCAACACAGTTGGGGCCCAAAATCAGTTGTGA 708
Db 235 MetAlaGlyIleHisAlaGlyValValSerAspValLeuGlyGlyGlnIleSerValVal 254
QY 709 ATTAGTAAGGTATTCCTTATTGAAGTCTTTGGCTAACACGTCACATCTGTGGTG 768
Db 255 IleSerLysGlyThrProTyrTyrGluSerSerLeuAlaAsnValThrSerMetVal 274
QY 769 GGACACTTACTACAAGTCTTTTACATTTAAGACAAAGTGGATGTTATGGAACACTGGG 828
Db 275 GlyTyrLeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGly 294
QY 829 ATGGAGTCTGGTGTGATCGGGATCTCAATACACATCATCTGTCTGGAGTGGACT 888
Db 295 MetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluTrpThr 314
QY 889 GACCACACAGGCAAGAGACAGTGTGAACCCCAAAACCCAGCTGAAAAACCTGGGA 948
Db 315 AspHisMetGlyGlnGluAsnSerTrpLysProGluLysAlaArgLeuArgLysProGly 334
QY 949 CCGCCCTTGGCTGCTTTTGGCACTGATGAATACCAAGTGGTGTACAAATAGATTGAATAAG 1008
Db 335 ProProTrpAlaAlaPheAlaThrAspGluHisGlnTrpLeuGlnIleAspLeuAsnLys 354
QY 1009 GAAAGAAATACAGGCATTAACCACTGGATCCACCATGCTGGAGCACAACTACTAT 1068
Db 355 GluLysLysIleThrGlyIleValThrThrGlySerThrLeuIleGluHisAsnTyrTyr 374
QY 1069 GTGCTCCCTACAGAACTGTACAGTATGATGGGAGAAATGAGTGTGTACAGAGAG 1128
Db 375 ValSerAlaTyrArgValLeuTyrSerAspAspGlyGlnLysTrpThrValTyrArgGlu 394
QY 1129 CCTGGTCTGGAGAAAGATATTTCAAGGAAACAAAGATTATCACAGGATGTCGCT 1188
Db 395 ProGlyAlaAlaGlnAspLysIlePheGlnGlyAsnLysAspTyrHisLysAspValArg 414
QY 1189 AATRACTTTTGGCACCACAAATTTGACGTTTATTAAGTGAATCTTACCAATGGCAG 1248
Db 415 AsnAsnPheLeuProIleAlaArgPheIleArgValAsnProValGlnTrpGln 434
QY 1249 CAGAAATTTGCCATGAAATGGAGCTGCTCGGATGTCAGTTTATTCCTAAAGTTCGTCCT 1308
Db 435 GlnLysIleAlaMetLysValGluLeuLeuGlyCysGlnPheThrLeuLysGlyArgLeu 454
QY 1309 CCAAACTTACTCAACCTCCACCTCCCTCGGAACAGCAATGACCTCAAAACACTACAGCC 1368
Db 455 ProLysLeuThrGlnProProProProArgAsnSerAsnAsnLeuLysAsnThrVal 474
QY 1369 CCTCCAAAATAGCAAAAGTGTGTCGCCCAAAATTTACGCAACCACTACACCTCGCAGT 1428
Db 475 HisProLysLeu-----GlyArgAlaProLysPheThrGlnAlaLeuGlnProArgSer 492
QY 1429 AGCAATGAATTTCTGTCACAGACAGACAACAACACTGCCAGTCTGATATCAGAAATACT 1488
Db 493 ArgAsnAspLeuProLeuLeuProAlaGlnThrThrAlaThrProAspValLysAsnThr 512
QY 1489 ACCGTAACTCCAAATGTAAACCAAGATGATGCTGGCTGCTGCTGCTGCTGCTGCTGCTG 1548
Db 513 ThrValThrProSerValThrLysAspValAlaLeuAlaValLeuValProValLeu 532
QY 1549 GTCATGTCCTCACTACTCTCTCATATAGTGTGTGCTGGCACTGGAGAAACAGA 1608
Db 533 ValMetAlaLeuThrLeuIleLeuIleLeuValCysAlaThrHisTrpArgAsnArg 552

QY 1609 AAGAAAAAACTGAAGCACCTATGACTTACTTGACGGGCGAGCTTGGTGGAAA 1668
Db 553 LysLysLysAlaGluGlyThrTyrAspLeuProHisTrpAspArgAlaGlyTyrTrpLys 572
QY 1669 GGAATGAAGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAAACCCAGTTCGCTAT 1728
Db 573 GlyValLysGlnLeuLeuProAlaLysSerValGluHisGluGluThrProValArgTyr 592
QY 1729 AGCAGCAGCAAGTAAATCACCCTGAGTCCAGAGAAAGTCCACAGTCTGCAGGCTGAC 1788
Db 593 SerAsnSerGluValSerHisLeuSerProArgGluValThrThrValLeuGlnAlaAsp 612
QY 1789 TCTGCAGATGATGCTCAGCCACTGCTAGGAGGAATTTGTTGACACATCTCAAGATCT 1848
Db 613 SerAlaGluThrAlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSer 632
QY 1849 ACCTTTAAACCCAGAAAGAAAGACAGGCTATCAGACCTAGATCTTACAACCTCA 1908
Db 633 ThrPheLysProGluGluGlyLysGluAlaSerTyrAlaAspLeuAspProTyrAsnAla 652
QY 1909 CCAGGCGAGGAAGTTTATCATCTGCTGACCACTCCCAATTCAGGGGCGCTGAGTAT 1968
Db 653 ProValGlnGluValTyrHisAlaTyrAlaGluProLeuProValThrGlyProGluTyr 672
QY 1969 GCAACCCCAATCATCATGGCATGTCCAGGCACCCCAACCTTCAGTTGTCAGCCCTCC 2028
Db 673 AlaThrProIleValMetAspMetSerGlyHisSerThrAlaSerValGlyLeuProSer 692
QY 2029 ACATCCACTTTCAAGTACGGGAGAACCAACCTCCCTCCAGTGGGAATTTACAATACA 2088
Db 693 ThrSerThrPheArgThrAlaGlyAsnGlnProProAlaLeuValGlyThrTyrAsnThr 712
QY 2089 CTTCCTCCAGGCTGACAGTGTCTCAGCCAGCCAGCCAGCTATGATACCCGAAAGCT 2148
Db 713 LeuLeuSerArgThrAspSerCysSerSerGlyGlnAlaGlnTyrAspThrProLysGly 732
QY 2149 GGGAAAGCAGGCTCTACCTGCCAGAGCAATTTGGTGTACCAAGTGCACAGAGACACAA 2208
Db 733 GlyLysProAla---AlaAlaProGluGluLeuValTyrGlnValProGlnSerThrGln 751
QY 2209 GAAGTATCAGAGCAGGAAGGATGGGAATGTGATGTTTTTAAAGAAATCTT 2262
Db 752 GluAlaSerGlyAlaGlyArgAspGluLysPheAspAlaPheLysGluThrLeu 769
RESULT 4
Q912V3 ID Q912V3 PRELIMINARY; PRT; 769 AA.
AC Q912V3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
DE protein.
DE protein.
GN ESDN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR;
RX PubMed=11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT "ESDN", a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury.";
RL J. Biol. Chem. 276:34105-34114(2001).
CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AF387548; AAL30179.1; .
DR MGD; MGI:1920629; ESDN.

DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR00421; FA58_C.
 DR InterPro; IPR004043; LCC_L_dom.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF00754; F5_F8_type_C; 1.
 DR PROSITE; PS01180; CUB; 1.
 KW DNA-binding; Transcription regulation.
 SQ SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;

Alignment Scores:
 Pred. No.: 6.31e-281 Length: 769
 Score: 3405.50 Matches: 642
 Percent Similarity: 90.24% Conservative: 42
 Best Local Similarity: 84.70% Mismatches: 61
 Query Match: 82.20% Indels: 13
 DB: 11 Gaps: 5

US-10-060-830-1 (1-2280) x Q912V3 (1-769)

QY 4 GCCGCCGCCGCC-----TGGCGCGCGCTCCCTCTCCCGC-----TCCTCTCCCT 48
 DB 20 AlaAlaProAlaAlaThrGlyArgAlaAlaLeuProSerAlaGlyCysCysProLeuPro 39
 QY 49 CCTGCTCCAACTCCTCCTCTCTCCATGCTCTGTTCTCTCCCTCTCTCTCTCTCTCTCT 108
 DB 40 ProGlyArgAsnSerSerSerArgProArg-----LeuLeuLeuLeuLeu 54
 QY 109 CTCCTGCTCGCTCGAGCGCTGGAGCCCGCAAGGTGATGGATGGACACACTGTACTA 168
 DB 55 LeuLeuLeuLeuGlnAspAlaGlyGlyGlnGlnGlyAspGlyCysGlyHisThrValLeu 74
 QY 169 GGCCCTGAGGTGGAACTTACATCCATAAAGTACACACAGACCTATCCCAACAGCACT 228
 DB 75 GlyProGluSerGlyThrLeuThrSerIleAsnTyrProHisThrTyrProAsnSerThr 94
 QY 229 GTTCTGTGATGGAGATCCGCTGTANAGATGGAGAGAGATTCGCATCAAAATTTGGTGAC 288
 DB 95 ValCysGluTrpGluLeuArgValArgThrGlyGluArgIleArgIleLysPheGlyAsp 114
 QY 289 TTTGACATTGAAGATCTTGATCTGTGTCACATTAACTTACTTGAGAATTTATAATGGAATT 348
 DB 115 PheAspIleGluAspSerAspTyrCysHisLeuAsnTyrLeuLysIlePheAsnGlyIle 134
 QY 349 GGAGTCAGCAGAACTGAAATAGGCAAACTACTGTGCTGCGGTGGCAATGAACATTCAC 408
 DB 135 GlyValSerArgThrGluIleGlyLysTyrCysGlyLeuGlyLeuGlnMetAsnGlnSer 154
 QY 409 ATTGATCAAAAGGCAATGAATACATTCATGCTGTCTCATGAGTGGNATCCATGTTCTCGA 468
 DB 155 IleGluSerLysGlySerGluValThrValLeuPheMetSerGlyThrHisAlaAlaGly 174
 QY 469 CGCGGATTTTGGCTCATCTCTCTTATAGATAAACAAGATCTAATTAATCTGTTTGGAC 528
 DB 175 ArgGlyPheLeuAlaSerTyrSerValIleAspLysGluAspLeuIleThrCysLeuAsp 194
 QY 529 ACTGCATCCAAATTTTGGAACTGAGTCACTAGTACTGCCAGCTGGTGTGCTGCT 588
 DB 195 ThrValSerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeu 214
 QY 589 CTTTTGCTGAGATATCTGGACAAATTCCTCATGATATAGATTCCTCGCCATTGGC 648
 DB 215 ProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerSerProLeuCys 234
 QY 649 ATGGCTGGTGTGCATGCAGGAGTACTGTCAAAACAGTTGGCGGCGCAATCAGTGTGTA 708
 DB 235 MetAlaGlyIleHisAlaGlyValValSerAsnValLeuGlyGlyGlnIleSerIleVal 254
 QY 709 ATTAGTAAAGGTATTCCTTATTATGAAAGTCTTTGGTCAACAGCTCACATCTGTGCT 768
 DB 255 IleSerLysGlyThrProTyrTyrGluSerSerLeuAlaAsnAsnValThrSerThrVal 274
 QY 769 GGACACTTATCTACAGCTCTTTTACATTTAAGACAAAGTGGATGTTATGGAACTGGGG 828
 DB 633 ThrPheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSer 552

DB 275 GlyTyrLeuSerAlaSerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGly 294
 QY 829 ATGGAGTCTGGTGTGATCGCGGATCCCTCAAAATAACAGCATCATCTGCTCGATGGAGCT 888
 DB 295 MetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerAlaLeuGluTrpThr 314
 QY 889 GACACACAGGCAAGAGACAGTGGAAACCCAAAAGCCAGCGCTCAAAAACCTCGA 948
 DB 315 AspHisMetGlyGlnGluAsnSerTrpThrAlaGluLysAlaArgLeuArgLysProGly 334
 QY 949 CCGCTTTGGCTGCTTTTGGCCACTGATGAATACCAGTGGTTACAAATAGATTGAATAG 1008
 DB 335 ProProTrpAlaAlaPheAlaThrAspGluHisGlnTrpLeuGlnIleAspLeuAsnLys 354
 QY 1009 GAAAGAAATACAGGCAATTAACCACTGGATCCCACTGGTGGGACACAAATACTAT 1068
 DB 355 GluLysLysIleThrGlyIleValThrThrGlySerThrMetIleGluHisSerTyr 374
 QY 1069 GTGTCTGCCTACACAATCCTGCTACAGTATGATGGCAGAAATGGACTGTGTACAGAG 1128
 DB 375 ValSerAlaTyrArgValLeuTyrSerAspAspGlyGlnArgTrpThrValTyrArgGlu 394
 QY 1129 CCTGCTGTGGAGCAAGATAAGATATTTCAAGGAAACAAAGATTATCACAGGATGTGGT 1188
 DB 395 ProGlyValAspGlnAspLysIlePheGlnGlyAsnLysAspTyrHisLysAspValArg 414
 QY 1189 AATAACTTTTGGCCACCAATTTATGACGTTTATTAGAGTGAATCCTACCACTGGAG 1248
 DB 415 AsnAsnPheLeuProIleIleAlaArgPheIleArgValAsnProValGlnTrpGln 434
 QY 1249 CAGAAATTTGCCATGAATGGAGCTGCTCGATGCTCAGTTTATTCCTTAAAGTGTCTCT 1308
 DB 435 GlnLysIleAlaMetLysValGluLeuLeuGlyCysGlnPheThrLeuLysGlyArgLeu 454
 QY 1309 CCAAACTTTACTCAACCTCCACCTCCTCGAACCACTACCTCAAAACACTACAGCC 1368
 DB 455 ProLysLeuThr-----ProProProArgAsnGlyAsnAsnLeuArgAsnThrAla 472
 QY 1369 CCTCCAAAATPAGCAAGTGTGTCGCCCAAAATTTACGCAACCACTACAACTCGCAGT 1428
 DB 473 ArgProLysLeuGlyLysGlyArgAlaProLysPheThrGlnValLeuGlnProArgSer 492
 QY 1429 AGCAATGAATTTCTGTCACACAGACAGAAACAACTGCCGCTGATATCAGAAATACT 1488
 DB 493 ArgAsnGluLeuProValGlnProAlaGluThrThrThrProAspIleLysAsnThr 512
 QY 1489 ACCGTAACTCCAAATGTAAACCAAGATGTAGCGCTGGCTGCAAGTTCTTGTCTCTGTGTG 1548
 DB 513 ThrValThrProSerValThrLysAspValAlaLeuAlaValLeuValProValLeu 532
 QY 1549 GTCATGTGCTCTACTCTCTCATATTATAGTGTGTGTGTGCTGGCACTGGAGAACAGA 1608
 DB 533 ValMetAlaLeuThrThrLeuIleLeuIleLeuValCysAlaTrpHisTrpArgAsnArg 552
 QY 1609 AAGAAAAAACTGAAGGCACTATGCTACTTACTGGGCGGCGGAGTTGGTGGAA 1668
 DB 553 LysLysLysThrGluGlyAlaTyrAspLeuProHisTrpAspArgAlaGlyTrpLys 572
 QY 1669 GGAATGAAGCAGTCTTCTTCCGCAAAAGCAGTGGACCACTGAGAAACCCAGCTTCGCTAT 1728
 DB 573 GlyMetLysGlnLeuLeuProAlaLysSerValAspHisGluGluThrProValArgTyr 592
 QY 1729 AGCAGCAGCGAAGTTAATCACCCTGAGTCCCAAGAGAGTCAACACAGTGTCTGAGGCTGAC 1788
 DB 593 SerThrSerGluValSerHisLeuSerAlaArgGluValThrThrValLeuGlnAlaAsp 612
 QY 1789 TCTGAGAGTATGCTCAGCCACTGCTAGAGGAATTTGGTGTACACTTCATCAAGATCT 1848
 DB 613 SerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSer 632
 QY 1849 ACCTTTAAACAGAGAAAGGAAAGCAAGCGCTATGACAGCTAGTACTCTTACAACTCA 1908
 DB 633 ThrPheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSer 552

QY 1909 CCAGGCGAGGAAGTTTATCATGCTGCTAGCTGAACCACTCCCAATTACGGGCGCTGAGTAT 1968
 Db 653 PrometGlnGluValTyrHisAlaTyrAlaGluProLeuProValThrGlyProGluTyr 672
 QY 1969 GCAACCCCATCATCATGAGCATGCTAGGAGGACCCCAACATTCAGTTGGTCAGCCCTCC 2028
 Db 673 AlaThrProIleValMetAspMetSerGlyHisProThrAlaSerValGlyLeuProSer 692
 QY 2029 ACATCCACTTTCAAGGCTACAGGGGAACCAACCTCCCTCCCTAGTGGGAACCTTACAATACA 2088
 Db 693 ThrSerThrPheLysThrAlaGlyThrGlnProHisAlaLeuValGlyThrTyrAsnThr 712
 QY 2089 CTTCTCTCCAGGAGTACAGCTGCTCTAGCCGCCAGCTAGTATGATPACCCGGAAGCT 2148
 Db 713 LeuLeuSerArgThrAspSerCysSerSerGlyGlnAlaGlnTyrAspThrProLysGly 732
 QY 2149 GGGAGCCAGGCTACCTGCTGCTAGCCGCCAGCTAGTGTACAGGTGCCACAGACACACAA 2208
 Db 733 GlyLys---SerAlaAlaThrProGluGluLeuValTyrGlnValProGlnSerThrGln 751
 QY 2209 GAAGTATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2262
 Db 752 GluLeuSerGlyAlaGlyArgAspGluLysPheAspAlaPheLysGluIleLeu 769

RESULT 5
 Q14089 PRELIMINARY; PRT; 364 AA.
 ID Q14089 AC Q14089
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 40.0 kDa protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shibata T.;
 RT "unpublished";
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL; D29810; BAA18909.1; -;
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR004043; LCCL_dom.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF00754; F5_F8_type_C; 1.
 DR SMART; SM00042; CUB; 1.
 DR PROSITE; PS01180; CUB; 1.
 KW Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;

Alignment Scores:
 Pred. No.: 1.59e-129 Length: 364
 Score: 1625.00 Matches: 342
 Percent Similarity: 94.81% Conservative: 5
 Best Local Similarity: 93.44% Mismatches: 14
 Query Match: 39.22% Indels: 11
 DB: 4 Gaps: 2

US-10-060-830-1 (1-2280) x Q14089 (1-364)

QY 145 GATGATGTGGACACACTGTACTAGCCCTGAGAGTGGAGCCCTTACATCCATAAATAC 204
 Db 1 AspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerIleAsnTyr 20
 QY 205 CCAGACACTTCCCAACAGCACTGTTGTGAATGGGAGATCCGTAAAGATGGGAGAG 264
 Db 21 ProGlnThrTyrProAsnSerThrValCysGluTyrGluIleArgValTysMetGlyGlu 40

QY 265 AGATTCCGATCAAAATTTGGTGACTTTGACATTGAAGATTCTGATTTTGTCTACTTTAAT 324
 Db 41 ArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCysHisPheAsn 60
 QY 325 TACTTGAAGATTTTATATGGAATTTGGAGTACGACAGACTGAAATAGCAATACTGTGGT 384
 Db 61 TyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLysTyrCysGly 80
 QY 385 CTGGGTTGCAATCAACCACTTCAATGAATCAAAAGCAATGAATCAACATCTCTGCTTC 444
 Db 81 LeuGlyLeuGlnIleAsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPhe 100
 QY 445 ATGAGTGAATCCATGTTCTGGACGCGGATTTTGGCCCTCATACTCTGTTATAGATAAA 504
 Db 101 MetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLys 120
 QY 505 CA-AGATCTAATTAATTTGGGACACTGCATCCAAAT-TTTTGGAACTGAGTTCAGTA 562
 Db 121 GlnArgSerAsnTyrLeuPheGlyHisCysIleGlnPhePheTyrPheAsnLeuSerVal 140
 QY 563 AGTACTGC-CCAGCTGGTGTCTGCTCTCTCTTCTGCTGAGATATCTGGAACAATTCCTCAT 621
 Db 141 SerThrAlaProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIleProHis 160
 QY 622 GGATATAGAGATTCCTGCCATTTGTCATGCTGGTGTGATGAGGAGTAGTGTCAAC 681
 Db 161 GlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValValSerAsn 180
 QY 682 ACCTGGGCGGCCAAATCACTGTTGTAATTAGTAAGGTATTCCTCTATTATGAAGTCT 741
 Db 181 ThrLeuGlyGlyGlnIleSerValIleSerLysGlyIleProTyrTyrGluSerSer 200
 QY 742 TTGGTAAACAACGTCACATCTGTGGTGGACACTTATCTACAAGTCTT---TTTACATTT 798
 Db 201 LeuAlaAsnValThrSerValValGlyHisLeuIleLeuGlnValPhePheThrPhe 220
 QY 799 AAGCAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGT-GATCGCGGATCCCA 857
 Db 221 LysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyGlyAspArgGlySerSer 240
 QY 858 AATAACACCATCATCTGTGCTGGAGTGGACTGACACACAGGCGCAAGAGACAGTTGAA 917
 Db 241 AsnAsnSerIle-ThrValLeuGluTyrThrAspHisThrGlyGlnGluAsnSerTyrPly 260
 QY 918 ACCCAAAAAGC-CAGGCTGAAAAAACCTGGACGCCCTGGGCTGCTTTTGGCCACTGATG 976
 Db 260 sProLysLysSerGlnAlaGluLysThrTyrThrAlaLeuGly-AlaPheAlaThrAspG 280
 QY 977 AATACCACTGGTTACAAATAGATTGTAATAAGGAAAAAATAACAGGCATTTATAACA 1036
 Db 280 IuTyrGlnTyrLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleThrThr 300
 QY 1037 CTGATCCACCATGGTGGAGCACAATTAATGCTGCTGCTCTACAGATCTCTGACATG 1096
 Db 300 hrGlySerThrMetVal-SerThrIleThrMetCysLeuProThrGluSerCysThrVal 319
 QY 1097 ATGATGGCAGAAATGGAATGTGTACAGAGACCTGGT-GTGGAGCAAGATAAGATATTT 1155
 Db 320 MetMetGlyArgAsnGlyLeuCysThrGluSerLeuValValGluGlnAspLysIlePhe 339
 QY 1156 CAAGGAAAAAAGAT---TATCACCAGGATGTCGTATAAATATTTTCCCAACCAATATT 1212
 Db 340 GlnGlyAsnLysArgIleIleThrArgMetValArgAsnAsnPheLeuProIleIle 359

QY 1213 GCACGTTTTATT 1224
 Db 360 AlaArgLeuLeu 363

RESULT 6
 Q9D9K5 ID Q9D9K5 PRELIMINARY; PRT; 251 AA.
 AC Q9D9K5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)

	1198	DB	:: :::	CAYCARMGNATHGCGNWTNAAGTNGARVNTNATHGNTGYCAR-----	1239
	335	QY	ProProLysLeuThrGlnProProProArgAsnSerAsnSp- :::	-----Leu	350
	1240	DB	-----ATHACNCAR-	-GGNAAYGAYWSNYTGNTGTGMGN	1272
	351	QY	LysAsnThrAlaProProLysIleAlaLysGlyArgAlaProLysPheThrGlnPro 	370	
	1273	DB	AARACNSWCARSWSNACNSWGNTWMSNACNAARAARGARGAYGARACNATHACNMGCNC 	1332	
	371	QY	LeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerPro :::	390	
	1333	DB	ATH- :::	--CCNWSNGARGACRACNSWNACNGN--	1359
	391	QY	AspIleArgAsnThrValThrProAsnValThrLysAspValAlaLeuAlaVal :	410	
	1360	DB	-----	---ATHAAYATHACNACNGTN	1377
	411	QY	LeuValProValLeuValMetValLeuThrThrLeuLeuLeuValCysAlaTrp ::: :	430	
	1378	DB	GCAHCHCCNWTNGTWNTYTNGTNTNTTYCCNGNGNATGGNATHYYCGMGCN ::: :	1437	
	431	QY	HisTrpArgAsnArgLysLysThrGluGlyThrTyraSPLeuProTyrrTpAspArg 	450	
	1438	DB	TTY-----MGNAARAAARAAARAAARGGNWSCNTAYGGNWGNCGARGCNCARAA :::	1491	
	451	QY	AlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGlu 	470	
	1492	DB	ACNGAYTGTTGGAAACARATHAAR-----TAYCCNTTYCGMGCNCAYCARWSNGCNGAR ::: :	1545	
	471	QY	ThrProValArgTyrrSerSerSerGluValAsnHisLeuSerProArgGluValThr :::	490	
	1546	DB	TTYACNATHWSNTAYGAY-----AAYGARAARGARATGACNCARAAARYTNGAY :::	1593	
	491	QY	ValLeuGlnAlaAspSerAlaGluTyrrAlaGlnProLeuValGlyGlyIleValGlyThr :::	510	
	1594	DB	YTNAHACNSWGNGAYATGCCNGAYTAYCARCCNTYNATGATGHGNCAGNCNGNAGTN :::	1653	
	511	QY	LeuHisGlnArgSerThrPhelysPro-----GluGluGlyLysGluAlaGlyTyrrAla :::	528	
	1654	DB	ACNWGAARGGWSNACNTTYMGCNCTGGAYCNGAYCGCARGARGCNGGN--GTN :::	1710	
	529	QY	AspLeuAspPro-----TyraSnSerProGlyGlnGluValTyrrHisAlaTyrrAla 	545	
	1711	DB	WSNACNGAYCGGNGCNCAVTAYGAYTYGCCARMGNCNGMGNCAYGARTAYGCN :::	1770	
	546	QY	GluProLeuProIleThrGlyProGluTyrrAlaThrProIleIleMetAspMetSerGly 	565	
	1771	DB	YTCCNTTYGNCNCNCNGARCCNGARAYGCNACCNCNATHGNTN-----GARMGN :	1821	
	566	QY	HisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGln :	585	
	1822	DB	CAYGTWYTMGNCNCAYACNTTYWSNCGNCARWSNGGNTAYMGNTGTCNGGCCNCAR :::	1881	
	586	QY	Pro-----ProProLeu-----ValGlyThr 	592	
	1882	DB	CCGNGNCAYAARCAYWSNTYTNWSNWSNGGNGGNTTYWSNCCGNTGCGNGGTNGGNCN :::	1941	
	593	QY	TyraSnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrrAspThr :::	612	
	1942	DB	CARGAYGGNGAYTAYCARMGNCNCAYWSNCGNCARCCNGCNGAYMGNGGNTAYCATYMGN :::	2001	
	613	QY	ProLysAla-----GlyLysProGlyLeuProAlaPro 	623	
	2002	DB	CCNAARGCNTNWSNGCNTTNGCNCARGWSNGNGCAYCCNGAYWSNCARAAARCCN 	2058	

```

: APPLICANT: Gao, Zeren
: APPLICANT: Shoemaker, Kimberly E.
: TITLE OF INVENTION: NEUROFILIN HOMOLOG 2CUB5
: FILE REFERENCE: 00-62
: CURRENT APPLICATION NUMBER: US/10/003,132
: CURRENT FILING DATE: 2001-11-15
: PRIOR APPLICATION NUMBER: US 60/249,004
: PRIOR FILING DATE: 2000-11-15
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 2836
: TYPE: DNA
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (129)...(1640)
US-10-003-132-3

Alignment Scores:
Pred. No.: 1,93e-29 Length: 2836
Score: 417.00 Matches: 147
Percent Similarity: 32.88% Conservatives: 70
Best Local Similarity: 22.27% Mismatches: 149
Query Match: 12.02% Indels: 294
DB: 9 Gaps: 18

US-10-060-830-1114 (1-653) x US-10-003-132-3 (1-2836)

Qy 13 LeuLeuLeuGluAspAlaGlyAlaGlnGlnGlyLysTyrCysGlyLeuGlyLeuGlnMet 32
Db 396 CTCCTCTTCAGCAGTGCAACAGATCAGATGATGGTCCATATTGTGGG--AGTTGGCGTGTT 452
Qy 33 AsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHis 52
Db 453 CCCAAGAAGCTCCGGCTCAACTCAACGAAGTAGCTGCTCTTCAAGAGTGATGATCTCAC 512
Qy 53 ValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThr 72
Db 513 ATTCTGGCGGGGCTTCTGTCGACCTACGCCAGCAGTAGCATCCCATTTAATAAC 572
Qy 73 CysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGly 92
Db 573 TGTTTGGAAAGCAGCAGCCATTATTTCGAGGAAAAATACAGCAAAATTCGCCCAGCTGGC 632
Qy 93 CysLeuLeuProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgaspSerSer 112
Db 633 TGTAGACATAGCAGGAGATATTTC7GGGAATACAAAAGATGGTTACAGAGATACCTCT 692
Qy 113 ProLeuCysMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIle 132
Db 693 TTATTGTCAAGCTGCCATCCAGCAGGAGATCATCAGATGAATAGTAGTGCCCAATC 752
Qy 133 SerValValIleSerLysGlyIleProTyrGluSerSerLeuAlaAsnAsnValThr 152
Db 753 AACTTGCTTCAGACGAAGGATAAGTCACATATGAAGGATCTCTGCGCAATGGCGTGCTC 812
Qy 153 SerValValGlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGly 172
Db 813 TCCCAGCATGTCTTCTTCGGAAGAGCATTTCTTTTACAAACC----- 857
Qy 173 ThrLeuGlyMetGlySerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeu 192
Db 857 ----- 857
Qy 193 GluTrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLys 212
Db 857 ----- 857
Qy 213 LysProGlyProProTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAsp 232
Db 857 ----- 857

```

QY 233 LeuAsnLysGluLysLysIleThrGlyIleIleThrThrGlySerThrMetValGluHis 252
Db 857 ----- 857
QY 253 AsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTrpThrVal 272
Db 857 ----- 857
QY 273 TyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyrHisGln 292
Db 858 ----- 858
QY 293 AspValArgAsnAsnPheLeuProIleIleAlaArgPheIleArgValAsnProThr 312
Db 866 ----- 866
QY 313 GlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheProLys 332
Db 866 ----- 866
QY 333 GlyArgProProLysLeuThrGlnProProProProArgAsnSerAsnAspLeuLysAsn 352
Db 866 ----- 866
QY 353 ThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPheThrGlnProLeuGln 372
Db 866 ----- 866
QY 373 ProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerProAspIle 392
Db 866 ----- 866
QY 393 ArgAsnThrThrValThrProAsnValThrLysAspValAlaLeuAlaValLeuVal 412
Db 867 ----- 867
QY 413 ProValLeuValMetValLeuThrThrLeuIleLeuValCysAlaTrpHisTrp 432
Db 900 ----- 900
QY 433 ArgAsnArgLysLysLysThrGluGlyThrTyrAspLeuProTyr 448
Db 945 ----- 945
QY 449 AspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHis 468
Db 990 ----- 990
QY 469 GluGluThrProValArgTyrSerSerGlu-----ValAsnHisLeuSerProArg 486
Db 1029 ----- 1029
QY 487 GluValThr-----ThrValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeu 503
Db 1077 ----- 1077
QY 504 ValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysPro----- 519
Db 1137 ----- 1137
QY 520 ---GluGlu-----GlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsn 534
Db 1197 ----- 1197
QY 535 SerProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGlu 554
Db 1242 ----- 1242
QY 555 TyrAlaThrProIleIle-----MetAspMetSer 564
Db 1302 ----- 1302
QY 565 GlyHisProThrThrSerValGlyGlnProSerThrThrPheLysAlaThrGlyAsn 584

Db 1362 GGCTACCGAGTCCCTGGG-----CCAGGCCACTCACAAACACTCCCATCTCTGGAGGC 1418
QY 585 GlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSer 604
Db 1419 TTTCTCTCTCTACAGGAGCCAGGTTGAAAGCTATCAGAGGCCAGGAGCCCAAG 1478
QY 605 SerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAsp 624
Db 1479 CCTGTGGTGGTGGCTATGACAGCCT---GCTGTAGCAGCTTCTTGGACAGCAGAGC 1535
QY 625 GluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGly 644
Db 1536 -----CCAGCCTCTCAGTCACAGATGACTCCGGGGGAGATGATGGT 1577
RESULT 8
US-09-823-038A-43
; Sequence 43, Application US/09823038A
; Patent No. US20020058335A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevin
; APPLICANT: Onrust, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Murison, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11090.1037c3
; CURRENT APPLICATION NUMBER: US/09/823,038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 1871
; TYPE: DNA
; ORGANISM: Mouse
US-09-823-038A-43
Alignment Scores:
Pred. No.: 1.3e-29 Length: 1871
Score: 416.00 Matches: 147
Percent Similarity: 32.88% Conservative: 70
Best Local Similarity: 22.27% Mismatches: 149
Query Match: 11.99% Indels: 294
DB: 10 Gaps: 18
US-10-060-830-1114 (1-653) x US-09-823-038A-43 (1-1871)
QY 13 LeuLeuLeuGluAspAlaGlyAlaGlnGlnGlyLysTyrCysGlyLeuGlnMet 32
Db 331 CTCTCTTTCAGCAGTGCACAGATCATGTGTCATATGTGGG---AGTTGGCTGT 387
QY 33 AsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHis 52
Db 388 CCCAAGAACTCCGCTGAACCTCAACGAAGTGAAGTCTCTTCAAGAGTGGATCTCAC 447
QY 53 ValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeuThr 72
Db 448 ATTTCTGGCGGGGCTTCTGTGTCACCTAGCCAGCAGTGCACCATCCAGATTTAATACC 507
QY 73 CysLeuSerThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGly 92
Db 508 TGTTTGGAGCAGGAGCCCATTTTCAGGAGAAAATACAGCAATTTCTGCCAGCTGGC 567
QY 93 CysLeuProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSer 112
Db 568 TGTAGACATAGCAGCAGATATTTCTGGGAATAAAGATGTTTACAGAGATACCTCT 627
QY 113 ProLeuCysMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlnIle 132
Db 628 TTATTGTGCAAGCTGCCATCCAGGAGGATCATCAGATGAACATAGGTGGCCACATC 687


```

1292 TCAGGACATATACAGACTATCAGATCACTGCTCCAGCATCTTCAGAACGCTCAACAG 1351
198 ThrGlyGlnGluAsnSerTyrLysProLysLysAlaArgLeuLysLysProGlyPro--- 216
1352 -----GACATGTTCACTGGGAACCAAGAAAGCTCGCTGGACAAGCAAGCAAGTG 1405
217 ---ProTrrPalaAlaPheAlaThrAspGluTyrGlnTrrPleuGlnIleAspLeuAsnLys 235
1406 AATGCTCGACCTCTGGCCACCAATGACCATGACATGTTTACAGTGGATCTTCTGTT 1465
236 GluLysLysIleThrGlyIleIleThrThrGlySerThrMetValGluHisAsnTyrTyr 255
1466 CCACCAAGTAGTGGCTGGCATCATACACAGGAGCTAAAGATTTCGTGTCATGACAGTTT 1525
256 ValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTrrPThrValTyrArgGlu 275
1526 GTTGGCTCTACAACTGGCTTACGCAATGATGGAGAACACTGGACTGTATACCAAGAT 1585
276 ProGlyValGluClnAspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspValArg 295
1586 GAAAGCAAAAGAAAGATAAGCTTTTCCAGGGAATTTTGACAAATGACACTCACAGAAA 1645
296 AsnAsnPheLeuProIleIleAlaArgPheIleArgValAsnProThrGlnTrrP 315
1646 AATGTCATGACCTCCCACTATGTCAGACACATTAAGATCTTCTGCTGCTGCTGCTG 1705
316 GlnLysIleAlaMetLysMetGluLeuLeuGlyCys 327
1706 GGGAGGATCACATGGGCTCAGAGCTGCTGGGCTGC 1741

```

RESULT 10

```

US-09-880-107-2275
; Sequence 2275, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Owe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2275
; LENGTH: 6909
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M16967
US-09-880-107-2275

```

```

Alignment Scores:
Pred. No.: 6,17e-23 Length: 6909
Score: 353.50 Matches: 108
Percent Similarity: 44.20% Conservatives: 33
Best Local Similarity: 33.86% Mismatches: 116
Query Match: 10.19% Indels: 62
DB: Gaps: 11

US-10-060-830-1114 (1-653) x US-09-880-107-2275 (1-6909)

Qy 50 GlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAsp 69
Db 5862 GGCITCAGAGTTCTGGGTTA--CTGGGAGCCCGAGATTACCAAGATTAAACAATGGTGG 5918
Qy 70 LeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCys 89

```

```

5919 AACTTATATGCTTG-----GAGTGTAGAAAACCTGC 5951
Qy 90 ProAlaGlyCys-----LeuLeuProPheAlaGluIleSerGlyThrIleProHis 106
5952 AGCAGAATTGCTCTTAAACCTTGGATCCAGGTGGACATGCAAAAGGAAGT-----CAT 6005
Qy 107 GlyTyrArgAspSerSerProLeuCysMetAla-----Gly 118
6006 AATCAGAGGATCCAGACCCCAAGGTGCCAAACACTACCTGAAGTCTCTGTATACCACAGA 6065
Qy 119 Val-HisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValValIleSer-- 137
6066 GTTCTATGTAGCTTACAGTTCACACCATCACTCACTGGCAGATCTTCAAAAGGGAACGAC 6125
Qy 138 -LysGlyIleProTyrTyrGlnSerSerLeuAlaAsnAsnValThrSerValValGlyHI 157
6126 AAGGAATGTGATGATATTTTAAT-----GGCAATTTCAGATGCCTCTACAAATAAAGA 6176
Qy 157 S----- 157
6177 GAATCAGTTTGACCCACCTATTGCTGGCTAGATATATAGGATCTCTCCAACTCGAGCCTA 6236
Qy 158 -----LeuSerThrSerLeuPheThrPheLysThrSerGlyCystyrGlyTh 173
6237 TAACACACCTACCTTCGATTGCAAGTGTGTGAGGTAAATGATGTTCCACACC 6296
Qy 173 rLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGI 193
6297 CCTGGTATGGAATAATGGAAGATAGAAAACAAGCAATCACAGCTTCTTCGTTTAAGAA 6356
Qy 193 u-----TrrPThrAspHisThrGlyGlnGluAsnSerTrrPlysProLysLysAlaArgLe 211
6357 ATCTTGGTGGGAGATTAC-----TGGGAACCTTCCTGCTGCCCTCT 6398
Qy 211 uLysLysProGly-----ProTrrPalaAlaPheAlaThrAspGluTyrGlnTrrPLe 229
6399 GAATGCCAGGAGCTGTGATGCTCGCAAGCAAGCAAGCAACAATAAGCAGTGGCT 6458
Qy 229 uGlnIleAspLeuAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 249
6459 AGAATTGATCTACTCAAGATCAAGAAGATAACGCGCAATTAACACAGGCTCGCAAGTC 6518
Qy 249 tValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLys 269
6519 TCTGCTCTCGAAATGATGTAAAGAGCTATACCATCCACTACAGTGGAGGAGTGGGA 6578
Qy 269 sTrrPThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAs 289
6579 ATGGAACCATACAGGCTGAAATCCTCCATGCTGGACAGATTTCGAAGGAATACTAA 6638
Qy 289 pTyrHisGlnAspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgVa 309
6639 TACCAAGGACATGTGAAGAATTTTCAACCCCAATCATTTCCAGGTTTATCCGCTGT 6698
Qy 309 lAsnProThrGlnTrrPLeuGlnLysIleAlaMetLysMetGluLeuLeuGlyCys 327
6699 CATTCCTAAAACATGGAATCAAGTATTACACTTCGCTCGGAACCTTTTGGCTGT 6753

```

RESULT 11

```

US-10-115-563-13
; Sequence 13, Application US/10115563
; Publication No. US20030008307A1
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H
; APPLICANT: Greengard, Judith S
; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATI-
; AND COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; Patent Counsel

```

STREET: 10666 No. US20030008307Alth Torrey Pines Road, TPC 8

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/115,563

FILING DATE: 02-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/410,488

FILING DATE: 24-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fitting, Thomas

REGISTRATION NUMBER: 34,163

REFERENCE/DOCKET NUMBER: 449.0

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-554-2937

TELEFAX: 619-554-6312

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 6925 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 175..6765

OTHER INFORMATION: /product= "Factor V"

FEATURE:

NAME/KEY: sig_peptide

LOCATION: 91..174

FEATURE:

NAME/KEY: CDS

LOCATION: 91..6765

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1..8

FEATURE:

NAME/KEY: misc_feature

LOCATION: 6918..6925

OTHER INFORMATION: /standard_name= "EcoRI linker"

nucleotide sequence"

SEQUENCE DESCRIPTION: SEQ ID NO: 13:

US-10-115-563-13

Alignment Scores:

Pred. No.: 6,19e-23 Length: 6925

Score: 353.50 Matches: 108

Percent Similarity: 44.20% Conservative: 33

Best Local Similarity: 33.86% Mismatches: 116

Query Match: 10.19% Indels: 62

DB: 9 Gaps: 11

US-10-060-830-1114 (1-653) x US-10-115-563-13 (1-6925)

Qy 50 GlytHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAsp 69

Db 5862 GGCCTCAGAGTTCTGGGTTA---CTGGAGCCAGATACAGATTAAACATGGTGG 5918

Qy 70 LeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerTyrCys 89

Db 5919 ATCTATAATGCTTG-----GAGCTGTAGAAAACCTTGC 5951

Qy 90 ProAlaGlyCys-----LeuLeuProPheAlaGluIleSerGlyThrIleProHis 106

Db 5952 AGCAGAAATTCCTCTAAACCTTGATCCAGGTGGACATGCAAAAGGAAGT-----CAT 6005

Qy 107 GlyTyrArgAspSerSerProLeuCysMetaIa-----Gly 118

Db 6006 AATCAGAGGATCCAGACCCAGAGTGCCCAACACACTACCTGAAGTCTCTGTATACCACAGA 6065

Qy 119 Val-HisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValIleSer--- 137

Db 6066 GTTCTATAGCTTACAGTTCCACAGATCAACTGGCAGATCTTCAAAGGAAGAACAGCAC 6125

Qy 138 -LysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThrSerValValGlyH1 157

Db 6126 AAGGAATGTGATGATTTTAAT-----GGCAATTTCAGATGCCTCTACAATAAAGA 6176

Qy 157 s----- 157

Db 6177 GAATCAGTTTGACCCACCTATTGTGGCTAGATATATTAGGATCTCTCCAACCTCGAGCCTA 6236

Qy 158 -----LeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGlyTh 173

Db 6237 TAACAGACCTACCTTCGATTGGAACTGCAAGTTGTGAGGTAAATGGATGTTCCACACC 6296

Qy 173 rLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuG1 193

Db 6297 CTGGGTATGGAATGGAAGATAGAAAACAAGCAATCACAGCTTCTCTCTTTAAGNA 6356

Qy 193 u-----TrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLe 211

Db 6357 ATCTGTGGGGAGATTAC-----TGGGAACCTTCCGTGCCGTCT 6398

Qy 211 uLysLysProGly-----ProTrpTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLe 229

Db 6399 GAATGCCCGAGGACGTGTGAATGCCTGGCAAGCCAAAGCAACAATAAAGCAGTGGCT 6458

Qy 229 uGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThrGlySerThrWe 249

Db 6459 AGAATTGATCTACTCAAGATCAAGAAGTAACGGCAATTAATACACAGGCTGCAAGTC 6518

Qy 249 tValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLy 269

Db 6519 TCTGTCTCTGAAATGATGTAACAGAGCTATACCTCCATCAGTACAGGAGGAGTGA 6578

Qy 269 sTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAs 289

Db 6579 ATGGAACCATACACAGGCTGAATCCTCATGCTGGACAGATTTTGAAGGAATACTAA 6638

Qy 289 pTyrHisGlnAspValArgAsnAsnPheLeuProIleIleAlaArgPheIleArgVa 309

Db 6639 TACCAAGGACATGTGAAGAACTTTTCAACCCCAATCATTTCCAGGTTTATCCGTGT 6698

Qy 309 lasnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuGlyCys 327

Db 6699 CATTCTAAACATGGAATCAAAAGTATTACACTTCGCCCTGGAACCTCTTTGGCTGT 6753

RESULT 12

US-10-115-563-26

; Sequence 26, Application US/10115563

; Publication No. US20030008307A1

; GENERAL INFORMATION:

; APPLICANT: Griffin, John H

; Greengard, Judith S

; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN

; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTA

; AND COMPOSITIONS THEREOF

; NUMBER OF SEQUENCES: 28

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: The Scripps Research Institute, Office of

; Patent Counsel

; STREET: 10666 No. US20030008307Alth Torrey Pines Road, TPC 8

; CITY: La Jolla

; STATE: CA

; COUNTRY: USA

; ZIP: 92037

; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn. Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/115,563
FILING DATE: 02-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/410,488
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: 449.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 6925 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-10-115-563-26

Alignment Scores:
Pred. No.: 6.19e-23 Length: 6925
Score: 353.50 Matches: 108
Percent Similarity: 44.20% Conservative: 33
Best Local Similarity: 33.86% Mismatches: 116
Query Match: 10.19% Indels: 62
DB: 9 Gaps: 11

US-10-060-830-1114 (1-653) x US-10-115-563-26 (1-6925)

QY 50 GlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAsp 69
DB 5862 GGCTTCAGAGTTCTTGGGTTA--CTGGAGCCAGATTAGCAAGATTAAACAATGGTGG 5918
QY 70 LeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCys 89
DB 5919 ATCTTATATAGTCTG-----GAGTGTAGAAAAAAGTGC 5951
QY 90 ProAlaGlyCys-----LeuLeuProPheAlaGluIleSerGlyThrIleProHis 106
DB 5952 AGCAGATTTGCCTTAACCTTGGATCCAGGTGGACATCCAAAGGAAGT-----CAT 6005
QY 107 GlyTyrArgAspSerProLeuCysMetala-----Gly 118
DB 6006 AATCAGAGGATCCAGACCCCAAGGTGCCAAACACTACCTGAAGTCTGTATACACAGA 6065
QY 119 Val-HisAlaGlyValValSerAsnThrLeuGlyGlyIleSerValIleSer-- 137
DB 6066 GTTCTATGTAGCTTACAGTTCCACACAGATCACTGCGCATCTTCAAGGGGAACAGCAC 6125
QY 138 -LysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThrSerValValGlyHi 157
DB 6126 AAGAAATGTGATGATTTTAAAT-----GGCAATTCAGATCCCTCTACATAAAGA 6176
QY 157 s----- 157
DB 6177 GAATCAGTTTACCACCACCTATTGGCTAGATATATTAGGATCTCTCAACTCGAGCCTA 6236
QY 158 -----LeuSerThrSerLeuPheThrPheLeysThrSerGlyCysTyrGlyTh 173
DB 6237 TAACAGACCTTACCCTTCGATTTGGAACTGCAAGGTGTGGAGGTAATGATGGATTTCCACACC 6296
QY 173 rLeuGlyMetGlySerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGl 193
DB 6297 CCTGGGTATGGAAATGGAAGATAGAAACAGCAATACACCTTCTTCGTTTAAAGAA 6356

QY 193 u-----TirThrAspHisThrGlyGlnGluAsnSerTirPlyProLysLysAlaArgLe 211
DB 6357 ACTTGTGGTGGGAGATAC-----TGGGAACCTTCCTCGTCCCGCTCT 6398
QY 211 uLysLysProGly-----ProProTirPAlaAlaPheAlaThrAspGluTyrGlnTirPLe 229
DB 6399 GAATGCCAGGAGCGTGTGAATGCCTGGCAAGCAAGCAACAACAATAAGCAGTGGCT 6458
QY 229 uGlnIleAspLeuAsnLysLysLysIleThrGlyIleIleThrGlySerThrMe 249
DB 6459 AGAATTTGATCTACTCAAGATCAAGAAGATAAGCGCAATATTATAACAGCGCTGCAAGTC 6518
QY 249 tValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLy 269
DB 6519 TCTGTCTCTGAATGTATGAAGAGCTATACCATCCATCAGTACAGGAGGAGTGGGA 6578
QY 269 sTirThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAs 289
DB 6579 ATGAAACCATACAGCGCTGAAATCCCTCCATGGTGGACAAGATTTTCAAGGAATACTAA 6638
QY 289 pTyrHisGlnAspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgVa 309
DB 6639 TACCAAGGACATGTGAAGAACTTTTCAACCCCAATCATTTCCAGGTTTATCCGCTGT 6698
QY 309 lAsnProThrGlnTirPLeGlnLysIleAlaMetLysMetGluLeuLeuGlyCys 327
DB 6699 CATTCCTAAACATGGAATCAAGATATTACATTCGCTCGCACTTTTGCTGT 6753

RESULT 13

US-09-880-107-3020
; Sequence 3020, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3020
; LENGTH: 1270
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 S56151
US-09-880-107-3020

Alignment Scores:
Pred. No.: 2.62e-23 Length: 1270
Score: 346.00 Matches: 84
Percent Similarity: 47.97% Conservative: 34
Best Local Similarity: 34.15% Mismatches: 94
Query Match: 9.97% Indels: 35
DB: 10 Gaps: 5

US-10-060-830-1114 (1-653) x US-09-880-107-3020 (1-1270)

QY 158 LeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMetGlu 177
DB 145 CTGCGCTTTGAGCTACTGGGCTGTGAGCTGAACGATGCGCAATCCCTCGGCTGAAG 204
QY 178 SerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGlu---TirThrAsp 196
DB 205 AATAACAGCATCCCTGACACAGCAGTACAGGCTCCAGCAGCTACAGACCTTGGGGCTG 264

Db 712 ----- 712
Qy 316 GlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheileProLysGlyArgPro 335
Db 712 ----- 712
Qy 336 ProLysLeuThrGlnProProProArgAsnSerAsnAspLeuLysAsnThrThrAla 355
Db 712 ----- 712
Qy 356 ProProLysIleAlaLysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSer 375
Db 712 ----- 712
Qy 376 SerAsnGluPheProAlaGlnThrGluGlnThrAlaSerProAspIleArgAsnThr 395
Db 712 ----- 712
Qy 396 ThrValThrProAsnValThrLysAspValAlaLeuAlaValLeuValProValLeu 415
Db 713 ----- 713
Qy 416 ValMetValLeuThrThrLeuLeuLeuValCysAlaThrHisThrArgAsnArg 435
Db 752 GCCCTCTCTGACTGGAATGGGATCTTTGCAATCTGT-----AGA 793
Qy 436 LysLysLysThrGluGlyThrThrAspLeuProThr-----TrpAspArgAla 451
Db 794 AAGAGGAAAAAGAAAGAAAT-----CCATATGTGTGTCAGCTGACGCTCAGAAACA 844
Qy 452 GlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThr 471
Db 845 GCCTGTTGGAGCAGATTAATATCCCTTT----- 874
Qy 472 ProValArgTrpSerSerGlu-----ValAsnHisLeuSerProArgGluValThr 489
Db 875 ---GCCAGGCATCAGTCGCGGAATTTACCATCAGCTATGACAAATGAAAGAGATGACA 931
Qy 490 -----ThrValLeuGlnAlaAspSerAlaGluThrAlaGlnProLeuValGlyGly 506
Db 932 CAAAGTTGGATCTCACTAGTATGATGCGCAGATTTATCAGCAGCCTCTCATGTTGCC 991
Qy 507 IleValGlyThrLeuHisGlnArgSerThrPheLysPro-----GluGlu 521
Db 992 ACAGGCACATCGGAGAAAGGCTCTACCTTCGACCCATGACACAGACACTGAGGAG 1051
Qy 522 -----GlyLysGluAlaGlyThrAlaAspLeuAspProThrAsnSerProGly 537
Db 1052 GTCAGAGTGAACACTGAGGCGCAGCGGCAC-----TATGACTGCTCTCAC 1096
Qy 538 GlnGluValThrHisAlaThrAlaGluProLeuProIleThrGlyProGluThrAlaThr 557
Db 1097 CGCGCGCGCGCATGATGATGACGCACTGCTTTCAGCAGCTCAGAACCTGATGCCACA 1156
Qy 558 Profilele-----MetAspMetSerGlyHisPro 567
Db 1157 CCTATCTGGAGCGGCACCTGCTCGAGCTCACACCTTCTCCACAGAGCGGCTACCGA 1216
Qy 568 ThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProPro 587
Db 1217 GTCCCTCGG---CCAGCGCCCACTCACAACACTCCCATTCCTGAGGCTTCTCTCTCT 1273
Qy 588 ProLeuValGlyThrThrAsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGln 607
Db 1274 GCTACAGAGCCGCCAGGTGAAAGCTATCAGAGGCCACAGCCCAAGCCCTGTGGGT 1333
Qy 608 AlaGlnThrAspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuVal 627
Db 1334 GGTGGCTATGACAGCCT---GCTGCTAGCAGCTTCTTGACAGCAGCAGAC----- 1381
Qy 628 TyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGly 644
Db 1382 -----CCAGCCTCTCAGTCACAGATGACTTCCGGGGGAGATGATGCT 1423

RESULT 15

US-09-150-811-7
GENERAL INFORMATION:
APPLICANT: Connelly, Sheila
Kaleko, Michael
Smith, Theodore
TITLE OF INVENTION: Adenoviral Vectors for Treatment of Hemophilias
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olst
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/150,811
FILING DATE: 10-Sep-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/484,891
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: 08/218,335
FILING DATE: 25-MAR-1994
APPLICATION NUMBER: 08/074,920
FILING DATE: 10-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Olstein, Elliot M.
REGISTRATION NUMBER: 24,025
REFERENCE/DOCKET NUMBER: 271010-440
TELECOMMUNICATION INFORMATION:
TELEPHONE: 973-994-1700
TELEFAX: 973-994-1744
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-150-811-7

Alignment Scores:

Pred. No.:	3 04e-20	Length:	4629
Score:	322.00	Matches:	72
Percent Similarity:	50.72%	Conservative:	33
Best Local Similarity:	34.78%	Mismatches:	80
Query Match:	9.28%	Indels:	22
DB:	10	Gaps:	4

US-10-060-830-1114 (1-653) x US-09-150-811-7 (1-4629)

Qy	153	SerValValGlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGly	172
Db	3886	AGCATCGCAGCACCTCTTCGCATGGAGTGTGATGGCTGTGATTAATAGTTCGACGATG	3945
Qy	173	ThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeu	192
Db	3946	CCATTGGGAATGAGAGTAAGCAATATCAGATGCACAGATTACTGCTTCATCTACTTT	4005
Qy	193	GluTrpThrAspHisThrGlyGlnGluAsnSerTrpLysLysAlaArgLeuLys	212
Db	4006	-----ACCAATATGTTGCCACCTGGTCTCTTCAAAAGCTCGACTTCAC	4050
Qy	213	LysProGly-----ProProTrpAlaAlaPheAlaThrAspGluThrGlnTrpLeuGln	230
Db	4051	CTCAAGGAGGAGGATGATGCTCGAGACCTCAGTGAATAATCCAAAGAGTGGCTGCA	4110
Qy	231	IleAspLeuAsnLysGluLysLysIleThrGlyIleIleThrThrGlySerThrMetVal	250
Db	4111	GTGAGCTTCCAGGACACAATGAAAGTCACAGGAGTAATACTCAGGAGTAATAATCTCTG	4170
Qy	251	GluHisAsnTrpThrValSerAlaThrArgIleLeuThrSerAspGlyGlnLysTrp	270

```
Db 4171 CTTACCAGCATGTATGTGAAGGAGTTCCCTCATCTCCAGCAGTCAAGATGCCATCAGTGG 4230
QY 271 ThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyr 290
Db 4231 ACTCTCTTTTTCAGAAATGGC-----AAAGTAAAGGTTTTTCAGGGAAATCAAGACTCC 4284
QY 291 HisGlnAspValArgAsnAsnPheLeuProIleIleAlaArgPheIleArgValAsn 310
Db 4285 TTCACACCTGTGTGNACTCTCTAGACCCACCGTTACTGACTCGCTACCTTGAATTAC 4344
QY 311 ProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPhe--- 329
Db 4345 CCCCAGAGTTGGGTGCACCATGTCCTGAGGATGGCGTCTGGGCTGCCAGGCACAG 4404
QY 330 -----IleProLysGlyArgProProLys 337
Db 4405 GACCTCTACTGAGGTGGCCACTGCGACACCTGCCCTACCTCTCCCTCCTCAG 4464
QY 338 LeuThrGlnProProPro 344
Db 4465 CTCCAGGGCAGTCTCCCTCCC 4485
```

Search completed: January 21, 2003, 12:59:07
Job time : 98.1281 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 09:35:55 ; Search time 8.84608 seconds

(without alignments)
8817.835 Million cell updates/sec

Title: US-10-060-830-1113

Perfect score: 3559

Sequence: 1 atgcctctgtctctctgct.....tttttaagaatacctttga 1962

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 241982

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10060830/runat_16012003_091241_26903/app_query.fasta_1.6933
-DB=PublishedApplications_AA -QFMT=fastan -SURFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10060830 -CGN 1 1 29 @runat_16012003_091241_26903
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications_AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap:
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap:
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap:
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap:
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap:
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap:
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap:
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap:
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap:
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap:
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap:
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pap:
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap:
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	873.5	24.5	715	9	US-10-003-132-2
2	417	11.7	503	9	US-10-003-132-4
3	416	11.7	503	10	US-09-823-038A-51
4	347	9.7	2224	9	US-10-115-563-14

Pred. No.:	4.56e-68	Length:	715
Score:	873.50	Matches:	225
Percent Similarity:	51.17%	Conservative:	102
Best Local Similarity:	35.21%	Mismatches:	209
Query Match:	24.54%	Indels:	103

ALIGNMENTS

RESULT 1
US-10-003-132-2
; Sequence 2, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUBS
; FILE OF INVENTION: 00-62
; CURRENT APPLICATION NUMBER: US/10/003.132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-132-2

5	325	9.1	1431	12	US-10-095-718-4	Sequence 4, Appli
6	324	9.1	458	9	US-10-003-132-6	Sequence 6, Appli
7	316	8.9	1438	12	US-10-006-091-1	Sequence 1, Appli
8	316	8.9	1438	12	US-10-047-257-1	Sequence 1, Appli
9	316	8.9	1471	12	US-10-095-718-2	Sequence 2, Appli
10	316	8.9	2332	9	US-09-957-641-2	Sequence 2, Appli
11	298	8.4	931	12	US-10-104-440-4	Sequence 4, Appli
12	266.5	7.5	923	12	US-10-104-440-2	Sequence 2, Appli
13	222.5	6.3	756	9	US-09-992-598-62	Sequence 62, Appli
14	222.5	6.3	756	9	US-09-989-293A-62	Sequence 62, Appli
15	222.5	6.3	756	9	US-09-989-735-62	Sequence 62, Appli
16	222.5	6.3	756	9	US-09-990-444-62	Sequence 62, Appli
17	222.5	6.3	756	9	US-09-989-730-62	Sequence 62, Appli
18	222.5	6.3	756	9	US-09-990-436-62	Sequence 62, Appli
19	222.5	6.3	756	9	US-09-991-181-62	Sequence 62, Appli
20	222.5	6.3	756	9	US-09-993-687-62	Sequence 62, Appli
21	222.5	6.3	756	9	US-09-989-734-62	Sequence 62, Appli
22	222.5	6.3	756	9	US-10-028-072-392	Sequence 392, App
23	222.5	6.3	756	9	US-09-997-653-62	Sequence 62, Appli
24	222.5	6.3	756	10	US-09-989-723-62	Sequence 62, Appli
25	222.5	6.3	756	10	US-09-989-723-62	Sequence 62, Appli
26	222.5	6.3	756	10	US-09-989-279-62	Sequence 62, Appli
27	222.5	6.3	756	10	US-09-989-727-62	Sequence 62, Appli
28	222.5	6.3	756	10	US-09-989-731-62	Sequence 62, Appli
29	222.5	6.3	756	10	US-09-989-732-62	Sequence 62, Appli
30	222.5	6.3	756	10	US-09-991-073-62	Sequence 62, Appli
31	222.5	6.3	756	10	US-09-990-442-62	Sequence 62, Appli
32	222.5	6.3	756	10	US-09-991-163-62	Sequence 62, Appli
33	222.5	6.3	756	10	US-09-993-604-62	Sequence 62, Appli
34	222.5	6.3	756	10	US-09-990-456-62	Sequence 62, Appli
35	222.5	6.3	756	10	US-09-989-721-62	Sequence 62, Appli
36	222.5	6.3	764	10	US-09-900-518A-2	Sequence 2, Appli
37	205	5.8	734	9	US-10-174-590-458	Sequence 458, App
38	205	5.8	734	9	US-10-176-586-458	Sequence 458, App
39	205	5.8	734	12	US-10-052-586-458	Sequence 458, App
40	200.5	5.6	855	10	US-09-771-161A-196	Sequence 196, App
41	185	5.2	913	10	US-09-223-490-4	Sequence 4, Appli
42	183	5.1	520	10	US-09-771-161A-105	Sequence 105, App
43	182.5	5.1	250	10	US-09-770-643A-8	Sequence 8, Appli
44	182.5	5.1	279	10	US-09-770-643A-10	Sequence 10, Appli
45	182.5	5.1	534	10	US-09-770-643A-14	Sequence 14, Appli

DB: 9 Gaps: 22
US-10-060-830-1113 (1-1962) x US-10-003-132-2 (1-715)
Qy 37 CTGCTGCTGAGGAGCGTGGAGCCAGCAGCAAGAAATACCTGCTGGTGGGGTGCAGATG 96
Dy 99 LeuLeuPheThrSerSerAspGlnThrGlyProThrCysGly---SerMetThrVal 117
Qy 97 AACCATTCATTAATGAATCAAAAGGCAATACACATGATGCTGCTCATGAGTGAATCCAT 156
Dy 118 ProLysGluLeuLeuLeuAsnThrSerGluValThrValArgPheGluSerGlySerHis 137
Qy 157 GTTCTGCGAGCGGATTTTGGCTCATCTGTTATAGATAAACAAGATCTAAATTAAT 216
Dy 138 IleSerGlyArgGlyPheLeuLeuThrTyrAlaSerSerAspHisProAspLeuLeuThr 157
Qy 217 TGTGTTGACACTGCATCCAAATTTTGGACCTGAGTTCAGTACAGTACCTCCAGCTGGT 276
Dy 158 CysLeuGluArgAlaSerHisTyrLeuLysThrGluTyrSerLysPheCysProAlaGly 177
Qy 277 TGTCTGCTCTCTTTGCTGAGATCTCGAACAATTCCTCATGATATACAGATTCCTCG 336
Dy 178 CysArgAspValAlaGlyAspIleSerGlyAsnMetValAspGlyTyrArgAspThrSer 197
Qy 337 CCATTGTGCTGCTGCTGATGAGTGTGCAAGGATGTGCAACACAGTGTGGCGGCCAAATC 396
Dy 198 LeuLeuCysLysAlaAlaIleHisAlaGlyIleIleAlaAspGluLeuGlyGlyGlnIle 217
Qy 397 AGTGTGTAATTAAGTATTCCTATATATGAAGTCTTTGGCTACACAGTGCACA 456
Dy 218 SerValLeuGlnArgLysGlyIleSerArgTyrGluGlyIleLeuAlaAsnGlyValLeu 237
Qy 457 TCTGTGGTGACACTATCTACAACTCTTTTACATTTAAGACNAGTGTGTTATGGA 516
Dy 238 SerArgAspGlySerLeuSerAspLysArgPheLeuPheThrSerAsnGlyCysSerArg 257
Qy 517 ACATGTGGGATGGAGTGTGCTGATCGCGGATCTCAAAATTAACAGCATCATCTGCTG 576
Dy 258 SerLeuSerPheGlu-----ProAspGlyGlnIleArgAlaSerSerSerTyr 273
Qy 577 GAGTGGAGTACACACAGGCAAGACAGACAGTGGTGAACCCAAAGCCAGCTGAAA 636
Dy 274 GlnSerValAsnGlnSerGlyAspGlnValHisTyrSerProGlyGlnAlaArgLeuGln 293
Qy 637 AACCTGGACCGCTTGGCTGCTTTGCCACTGATGAA-----TACCAGTGG 684
Dy 294 AspGlnGlyProSerTyrPheSerGlyAspSerSerAsnAsnHisLysProArgGluTyr 313
Qy 685 TTCAAAATAGATTTGAATAAGAAAGAAATAACAGGCAATTTATACCACTGGATCCACC 744
Dy 314 LeuGlnIleAspLeuGlyGlyLysLysIleThrGlyIleArgThrThrGlySerThr 333
Qy 745 ATGGTGGAGCAATTAATCTGCTGCTCCATACAGATCTCTGATGATGATGGCAG 804
Dy 334 GlnSerAsnPheAsnPheTyrValLysSerPheValMetAsnPheLysAsnAsnSer 353
Qy 805 AAATGGAGCTGTGTACAGAGACCTGGTGTGAGCAAGATTAAGATTTTCAAGAAACAAA 864
Dy 354 LysTyrPheThrTyrLysGlyIleValAsnAsnGluGluLysValPheGlnGlyAsnSer 373
Qy 865 GATTATACAGGATGCGGTAATTAATCTTTTGGCCCAATTAATGACGCTTTTATTAGA 924
Dy 374 AsnPheArgAspProValGlnAsnAsnPheIleProIleValAlaArgTyrValArg 393
Qy 925 GTGAATCTTACCATGCGCAGCAAAATTCCTATGAAATGGAGCTGCTGGATGTCAG 984
Dy 394 ValValProGlnThrPheHisGlnArgIleAlaLeuLysValGluLeuIleGlyCysGln 413
Qy 985 TTTATTCTAAAGGTGCTGCTCCAAAACCTTACTCAACCTCCACCTCTCGGAACAGCAAT 1044
Dy 414 -----IleThrGln-----GlyAsn 418
Qy 1045 GAC-----CTCAAAACACTACAGCCCTCCAAAATAGCCAAAGGTGCTGCC 1092

DB 419 AspSerLeuValTyrArgLysThrSerGlnSerThrSerValSerThrLysLysGluAsp 438
Qy 1093 CCAAAATTTACGCAACCACTACAACCTCGCAGTAGCATGAATTTCTTCCGACACAGCAA 1152
Dy 439 GluThrIleThrArgProIle-----ProSerGlu 448
Qy 1153 CAAACCACTGCCAGTCTGATATCAGAAATACCTACCGTAACCTCCAAATGTAAACCAAGAT 1212
Dy 449 GluThr-----SerThrGlyIleAsnIleThr----- 458
Qy 1213 GTACGCTGGCTGAGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1272
Dy 459 ValAlaIleProLeuValLeuValLeuValLeuValLeuValLeuValLeuValLeuValLeu 478
Qy 1273 ATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1332
Dy 479 AlaPhe-----ArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 488
Qy 1333 TTACTCTTACTGG-----GACCGGCAGCTTGTGGAAGGAATGAAGCAGCTTT 1380
Dy 489 ---ProTyrGlySerAlaGluAlaGlnLysThrAspCysTyrLysGlnIleLysTyr--- 506
Qy 1381 CTTCCTGCAAAAGCAGTGGACCATGAGGAAACCCCA-----GTTCGCTATAGCAGCAGCAA 1437
Dy 507 -----ProPheAlaArgHisGlnSerAlaGlu 515
Qy 1438 -----GTTAATCACCCTGAGTCCAAAGAGAGTCACTCACC-----ACAGTCTGCTGAGCT 1482
Dy 516 PheThrIleSerTyrAspAsnGluLysGluMetThrGlnLysLeuAspLeuIleThrSer 535
Qy 1483 GACTCTCAGATGATGCTCAGCCTGCTGAGGAGGAATTTGTTGTCACACTTCATCAAGA 1542
Dy 536 AspMetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrValThrArgLysGly 555
Qy 1543 TCTACCTTTAAACCA-----GAAGAGAAAGAGAGAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1596
Dy 556 SerThrPheArgProMetAspThrAspAlaGluAlaGly---ValSerThrAspAla 574
Qy 1597 -----TACACTCAGCAGGCGAGGAGGAGTATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1647
Dy 575 GlyGlyHisTyrAspCysProGlnArgAlaGlyArgHisGluTyrAlaLeuProLeuAla 594
Qy 1648 ATTACGGGGCTGAGTATGCAACCCCAATCATCATGAGGAGTGTGTCAGGCGACCCCAACT 1707
Dy 595 ProProGluProGluTyrAlaThrProIleVal-----GluArgHisValLeuArg 611
Qy 1708 TCAGTGTGCTGAGCCCTCCATCCACTTTCAGGCTCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1758
Dy 612 AlaHisThrPheSerAlaGlnSerGlyTyrArgValProGlyProGlyHisLys 631
Qy 1759 -----CCCCCACTA-----GTGGGAACCTTACATACACTT 1788
Dy 632 HisSerLeuSerSerGlyPheSerProValAlaGlyValGlyAlaGlnAspGlyAsp 651
Qy 1789 CTCTCCAGGACTGACGCTGCTCCTCAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1845
Dy 652 TyrGlnArgProHisSerAlaGlnProAlaAspArgGlyTyrAspArgProLysAla 670

RESULT 2

US-10-003-132-4
: Sequence 4, Application US/10003132
: Publication No. US20020192750A1
: GENERAL INFORMATION:
: APPLICANT: Fox, Brian A.
: APPLICANT: Gao, Zeren
: APPLICANT: Shoemaker, Kimberly E.
: TITLE OF INVENTION: NEUROFILIN HOMOLOG 2CUB5
: FILE REFERENCE: 00-62
: CURRENT APPLICATION NUMBER: US/10/003,132
: PRIOR FILING DATE: 2001-11-15
: PRIOR FILING DATE: 2000-11-15

```
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-003-132-4

Alignment Scores:
Pred. No.: 2,38e-28 Length: 503
Score: 417.00 Matches: 148
Percent Similarity: 32.98% Conservative: 70
Best Local Similarity: 22.39% Mismatches: 147
Query Match: 11.72% Indels: 296
DB: 9 Gaps: 19

US-10-060-830-1113 (1-1962) x US-10-003-132-4 (1-503)

QY 37 CTGCTGCTCGAGGACCTGGAGCCAGCAAGCAATCTAGTGGTCTGGGGTGGCAATG 96
Db 90 LeuLeuPheSerSerAlaThrAspGlnTyrGlyProTyrCysGly---SerTrpAlaVal 108
QY 97 AACCAATCAATTGAATCAAAAGGCAATGAATCAATCATCTGTTTCATGAGTGGAAATCCAT 156
Db 109 ProLysGluLeuArgLeuAsnSerAsnGluValThrValLeuPheLysSerGlySerHis 128
QY 157 GTTCTGGACGGGATTTTGGCCCTCATCTCTGTATATAGATAAACAAGATCTAATTAAT 216
Db 129 IleSerGlyArgGlyPheLeuLeuThrTyrAlaSerSerAspHisProAspLeuIleThr 148
QY 217 TCTTTGGACATGCAATCAATTTTGGACCTGAGTTCAGTAAGTACTGCCCGAGTGT 276
Db 149 CysLeuGluArgGlySerHisTyrPheGluGluLysTyrSerLysPheCysProAlaGly 168
QY 277 TCTGCTGCTCTTTCGTGAGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTCG 336
Db 169 CysArgAspIleAlaGlyAspIleSerGlyAsnThrLysAspGlyTyrArgAspThrSer 188
QY 337 CCATTGTGCATGGCTGTGTCATGACGAGTAGTGTCAACACGCTTGGCGGCCCAATC 396
Db 189 LeuLeuCysLysAlaAlaIleHisAlaGlyIleIleThrAspGluLeuGlyGlyHisIle 208
QY 397 AGTGTGTAAATAGTAAGGTATTCCTATTATGTAAGTCTTTGCTTAACACGTCACA 456
Db 209 AsnLeuLeuGlnSerLysGlyIleSerHisTyrGluGlyLeuLeuAlaAsnGlyValLeu 228
QY 457 TCTGTGTGGGACACTATCTACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGA 516
Db 229 SerArgHisGlySerLeuSerGluLysArgPheLeuPheThrThr----- 243
QY 517 ACACCTGGGATGGAGTCTGTGTGTCGCGGATCCTCAATAACAGCATCATCTGTGCTG 576
Db 243 ----- 243
QY 577 GAGTGACTGACACACAGGGCAAGAGACAGTGTGGAACCCCAAAAGCCAGGTGAAA 636
Db 243 ----- 243
QY 637 AACCTGGACCGCTTGGGCTGCTTTGGCCACTGATGAATACACAGTGGTTACAATAGAT 696
Db 243 ----- 243
QY 697 TTGAATAAGGAAAGAAATAACAGGCAATATAACCACTGGATCCACCATGGTGGAGCAC 756
Db 243 ----- 243
QY 757 AATTACTATGTGTCTGCCTACAGAATCTGTATACAGTGTATGATGGCGCAAGAAATGAGTGTG 816
Db 243 ----- 243
QY 817 TACAGAGACGCTGGTGGAGCAAGATAGATATTTCAAGGAACAAGATATTACACAG 876
Db 244 -----ProglyMet----- 246
```

```
QY 877 GATGTCGCTAATAACTTTTTGCCACAATATTATTCACGTTTATTATAGACTGAATCCTACC 936
Db 246 ----- 246
QY 937 CAATGCCACGAGAAATTTGCCATGAATAATGGAGCTGCTCGGATGTCAGTTTATTCTCTAA 996
Db 246 ----- 246
QY 997 GGTGCTCTCCAAACTTACTCAACCTCCACCTCCTCGAAGCAAGCAATGACCTCAAAAAC 1056
Db 246 ----- 246
QY 1057 ACTACAGCCCTCCAAAAATAGCCAAAGGTGCTGCCCAAAATTTACGCAACCACTACAA 1116
Db 246 ----- 246
QY 1117 CCTCGCAGTAGCAATGAATTTCTGCACAGACAGAAACAACCTGCCAGTCTGTATATC 1176
Db 246 ----- 246
QY 1177 AGAATACTACCGTAACCTCCAAATGTAAACCAAGATGTAGCGTGGCTGCTGCTTTGTGTC 1236
Db 247 -----AsnIleThr---ThrValAlaIleProSerValIle--- 257
QY 1237 CCGTGTGCTGTCATGCTCTCTCACTACTCTCATTTCTATATTTAGTGTGCTTGGCACTGG 1296
Db 258 ---PheIleAlaLeuLeuLeuThrGlyMetGlyIlePheAlaIleCys----- 272
QY 1297 AGAAACAGAAAGAAAACCTGAAGGCACCTATGACTTACCTTAC-----TGG 1344
Db 273 -----ArgLysArgLysLysLysGlyAsn-----ProTyrValSerAlaAspAla 287
QY 1345 GACCGGCGAGTGTGGTGAAGGAATGAAGCAGTTTCTTCTGCCAAAAGCAGTGGACCAT 1404
Db 288 GlnLysThrGlyCysTrpLysGlnIleLysTyr----- 298
QY 1405 GAGGAACCCCA---GTTGCTATACGACGACGCA---GTTAATCACCTGAGTCCA 1455
Db 299 -----PropheAlaArgHisGlnSerThrGluPheThrIleSerTyrAspAsnGlu 315
QY 1456 AGAAGAGTCAAC-----ACAGTGTGTCAGGCTGCTGTCAGAGTATGTCAGCCA 1506
Db 316 LysGluMetThrGlnLysLeuAspLeuIleThrSerAspMetAlaAspTyrGlnGlnPro 335
QY 1507 CTGCTAGGAGGAATTTGTTGTACACTTCATCAAGATCATCTTTTAAACCA----- 1557
Db 336 LeuMetIleGlyThrGlyThrValAlaArgLysGlySerThrPheArgProMetAspThr 355
QY 1558 -----GAAGAA-----GGAAAGACGAGGCTATGTCAGACCTAGATCCTTAC 1599
Db 356 AspThrGluGluValArgValAsnThrGluAlaSerGlyHis-----Tyr 370
QY 1600 AACTCACCAGGCGAGGAGTTTATCATGCTCATGCTGAACCACTCCCAATTTACGGGCGCT 1659
Db 371 AspCysProHisArgProGlyArgHisGluTyrAlaLeuProLeuThrHisSerGluPro 390
QY 1660 GAGTATCAACCCCAATCATC-----ATGGACATG 1689
Db 391 GluTyrAlaThrProIleValGluArgHisLeuLeuArgAlaHisThrPheSerThrGln 410
QY 1690 TCAGGGACCCCAACAACCTCAGTTGGTTCAGCCCTCCACATCCCACTTTCAAGGCTACGGG 1749
Db 411 SerGlyTyrArgValProGly---ProArgProThrHisLysHisSerHisSerSerGly 429
QY 1750 AACCAACCTCCCACTAGTGGGAACTTACATACACTTCTCTCCAGGACTGACAGCTGC 1809
Db 430 GlyPheProAlaThrGlyAlaThrGlnValGluSerTyrGlnArgProAlaSerPro 449
QY 1810 TCCTCAGCCCGCCAGGTATGATACCCGAAAGCTGGGAAGCAGGTCTTACCTGCCCA 1869
Db 450 LysProValGlyGlyTyrAspLysPro---AlaAlaSerSerPheLeuAspSerArg 468
```



```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

Alignment Scores:
Pred. No.: 3.65e-20 Length: 1431
Score: 325.00 Matches: 99
Percent Similarity: 43.60% Conservative: 51
Best Local Similarity: 28.78% Mismatches: 121
Query Match: 9.13% Indels: 73
DB: 12 Gaps: 11

US-10-060-830-1113 (1-1962) x US-10-095-718-4 (1-1431)
QY 48 GGAGCTGGAGCCCA---GCAAGGCAATACTGTGGTCTGGGTTGCAAAAT----- 95
Db 1143 GlycylTrpAlaProLysLeuAlaArgLeuHisTyrSerGlySerIleAsnAlaTrpSer 1162
QY 96 -----GAACCAATCAATGAATCAAAAGCAATGAATCAATCAATGCTGTTCATGAGTGG 149
Db 1163 ThrLysAspProPheSerTrpIle-LysValAspLeuLeuAlaProMetIleHisG1 1182
QY 150 AATCCATGTTTCGGACGCGGATTTTGGCCTCATACTCTCTTATAGATAAACAAGATCT 209
Db 1182 YlleMetThrGlnGly----- 1187
QY 210 AATTACTGTTTGGACACTGCATCCAAATTTTGGAACTGAGTTCACTAGTACTGCC 269
Db 1188 -----AlaArgGlnLysPheSerLeuTyrVa 1197
QY 270 AGCTGGTGTCTGCTTCTCTTTGCTAGATATCTGGAAACAATCTCATGATATAGAGA 329
Db 1197 LserGlnPheIleIleMetTyrSer---LeuAspGlyAsnLysTrpHisSerTyrArgG1 1216
QY 330 TTCCTCCCATTTGTCATGGCTGTGTGTCATGCAGGAGTAGTGTCAAACACGTTGGCGG 389
Db 1216 yAsnSer-----ThrGlyThrLeuMetVa 1224
QY 390 CCAATCAGTGTGTAATAGTAAGTATT-----CCCTATTA 428
Db 1224 lPhePheGlyAsnValAspSerSerGlyIleLysHisAsnIlePheAsnProProlIle1 1244
QY 429 TGAAGTCTTGTGCTAACACAGCTACA-----TCGTGTGGTGGACACTTATCTACAAG 482
Db 1244 eAlaGlnTyrIleArgLeuHisProThrHisTyrSerIleArgSerThrLeuArgMetG1 1264
QY 483 TCTTTTACATTTAAGACAAGTGTATGGAACACTGGGATGGAGTCTGTGTGAT 542
Db 1264 uLeuLeuGlyCysAspPheAsnSerCysSerMetProLeuGlyMetGluSerLysAlaI1 1284
QY 543 CGCGGATCCTCAAAATAACAGCATCATCTGTGGTGGAGTGGACTGACCACACAGGCAAGA 602
Db 1284 eSerAspAlaGlnIleThrAlaSerSerTyrLeu-----SerSerMetLe 1299
QY 603 GAACAGTTGGAAACCCAAAAGCCAGCTGAAAAAACCTGGA-----CCGCTTGGGC 656
Db 1299 uAlaThrTrpSerProSerGlnAlaArgLeuHisLeuGlnGlyArgThrAsnAlaTrpAr 1319
QY 657 TGCCTTTGCCACTGATGATACAGTGGTTACAATAGATTGTAATAGGAAAGAAAT 716
Db 1319 gProGlnAlaAsnAsnProLysGluTrpLeuGlnValAspPheArgLysThrMetLysVa 1339
QY 717 AACAGGATTAATAACACTGCATCCACCATGGTGAGGACAAATTAATGCTGCTGCTCA 776
Db 1339 lThrGlyIleThrThrGlnGlyValLysSerLeuLeuIleSerMetTyrValLysGlnPh 1359
QY 777 CAGAACTCTGTACAGTATGATGGCAGAAATGGAATGGAATGCTGTACAGAGAGCTGTGTGGA 836
Db 1359 eLeuIleSerSerGlnAspGlyHisAsnTrpThrLeuPheLeuGlnAsnGly----- 1377
```

```
QY 837 GCAAGATAAGATATTTCAAGSAAACAAGATATTATCACCAGGATCTGCGTAATACTTTT 896
Db 1378 -LysValLysValPheGlnGlyAsnArgAspSerSerThrProValArgAsnArgLeuG1 1397
QY 897 GCCACCAATATTATTCACGCTTTTATTAGAGTGAATCTTACCAATGGCAGCAGAAAATTGC 956
Db 1397 uProProLeuValAlaArgTyrValArgLeuHisProGlnSerTrpAlaHisIleAl 1417
QY 957 CATGAAATGAGCTGCTCGATGTCAGTTTATCTTAAAGTCTGCTCTCCAAAACCTTAC 1016
Db 1417 aLeuArgLeuGluValLeuGlyCysAsp-----Th 1427
QY 1017 TCAACCTCCA 1026
Db 1427 rGlnGlnPro 1430

RESULT 6
US-10-003-132-6
; Sequence 6, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-003-132-6

Alignment Scores:
Pred. No.: 2.92e-20 Length: 458
Score: 324.00 Matches: 140
Percent Similarity: 31.09% Conservative: 77
Best Local Similarity: 20.06% Mismatches: 159
Query Match: 9.10% Indels: 322
DB: 9 Gaps: 22

US-10-060-830-1113 (1-1962) x US-10-003-132-6 (1-458)
QY 4 CCTCTGTTCTCTCTCTTACTTGTCTGCTGCTGCTGCTGAGGACGCTGGAGCCCGAG 63
Db 8 ProSerValLeuAlaLeuLeuPheAlaValCysAlaProLeuArgLeuGlnAlaGluG1 27
QY 64 CAAGGCAAACTACTGTGT----- 81
Db 28 LeuGlyAspGlyCysGlyHisIleValThrSerGlnAspSerGlyThrMetThrSerLys 47
QY 82 -----CTGGGTTGCAAAATGAACCATCAAT-----GAATCAAAA 117
Db 48 AsnTyrProGlyThrTyrProAsnTyrThrValCysGluLysIleIleThrValProLys 67
QY 118 GGCAATGAATCACAATCTCTTTCATGAGTGAATCCATGTTCTGCGAGCGGATTTTG 177
Db 68 GlyLysArgLeu-----IleLeuArgLeuGlyAspLeuAsnIleGluSerLysThrCysAla 86
QY 178 GCCTCATAC-----TCTGTTATAGATAAACAAGATCTAATTAATCTGTTGGAC 225
Db 87 SerAspTyrLeuLeuPheSerSerAlaThrAspGlnTyrAspLeuIleThrCysLeuGlu 106
QY 226 ACTGCATCCAAATTTTGGAAACCTGAGTTCAGTAAGTACTGCCAGCTGGTGTGTGCT 285
Db 107 ArgGlySerHisTyrPheGluGluLysTyrSerLysPheCysProAlaGlyCysArgasp 126
QY 286 CCTTTTCTGATGATATCTGGAAACAATTCCTCATGATATAGAGATTCCTGCCCATGTC 345
```

Alignment Scores:	
Pred. No.:	2,228-19
Score:	316.00
Percent Similarity:	44.4%
Best Local Similarity:	29.1%
Query Match:	8.8%
DB:	12
Gaps:	12
Indels:	69
Mismatches:	116
Conservative:	51
Matches:	97
Length:	1438

```
US-10-060-830-1113 (1-1962) x US-10-006-091-1 (1-1438)
Qy 48 GGAGCTGGAGCCCA---GCAGCAATAGTCTGGTCTGGGTGCAAT----- 95
Db 1150 GlyGlnTrpAlaProLysLeuAlaArgLeuHisThrSerGlySerIleAsnAlaTrpSer 1169
Qy 96 -----GAACATTCATTAATCAAAAGCAATGAATCAATGCTGCTTCATGAGTGG 149
Db 1170 ThrLysGluProPheSerTrpIle-LysValAspLeuAlaProMetIleIleHisG1 1189
Qy 150 AATCCATGTTCTCGA-----CGCGGATTTTGGC 179
Db 1189 yileLysThrGlnGlyAlaArgGlnLysPheSerSerLeuTyriLeSerGlnPheIleI1 1209
Qy 180 CTCATATCTCTTATAGTAAACAGATCTAATTAATCTGTTGTCGACACTGCATCCATTT 239
Db 1209 eMetTySerLeuAspGlyLys----- 1216
Qy 240 TTTGGAACCTGAGTTCAGTAAGTAC-----TGCCAGCTGGTGTCTGCTTCCTTTTGC 293
Db 1217 -----LysTrpGlnThrTyArgGlyAsnSerThrGlyThrLeuMetValPhe-- 1232
Qy 294 TGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTCGCCATTTGTCGATGGCTGG 353
Db 1233 -----PheGlyAsnValAspSerSer----- 1239
Qy 354 TGTGCATGCAGGAGTAGTGTCAACACACCTTGGCGGCGCAATCAGTGTGTGAATAGTAA 413
Db 1240 -----GlyIleLysHisAsnIlePheAsnProPheIle-----IleAlaAr 1253
Qy 414 AGGTATTCCTTATTAAGAACTCTTTTGGCTTAACAACGTCACATCTGTGGTGGGACACTT 473
Db 1253 gTyriLeArgLeuHisProThrHisTy-----SerIleArgSerThrLe 1268
Qy 474 ATCTACAGTCTTTTACATTTAACAAGTGAAGTGTATGAACACATGGGAGGAGTGC 533
Db 1268 uArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMetProLeuGlyMetGluSe 1288
Qy 534 TGGTGTGATCCGGATCCTCAATTAACAGCATCATCTGCTGCTGGAGTGGACACAC 593
Db 1288 rLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyPhe-----Th 1303
Qy 594 AGGCAAGAGAACAGTGTGAACCCAAAGCCAGGCTGAAAAAACCCTGGA-----CC 647
Db 1303 rAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHisLeuGlnGlyArgSerAs 1323
Qy 648 GCCTTGGCTGCTTTTGGCCACTGATGATACCACTGGTTCATTAAGATTTGAATAAGGA 707
Db 1323 nAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGlnValAspPheGlnLysTh 1343
Qy 708 AAGAAATACAGGCATTAACCACTGGATCCACCATGCTGGAGGACACAAATTAATGT 767
Db 1343 rMetLysValThrGlyValThrThrGlnGlyValLysSerLeuLeuThrSerMetTyRva 1363
Qy 768 GTCTGCCTACAGATCCTGTACAGTGTAGTGGGACAGAAATGGACTGTGTACAGAGACC 827
Db 1363 lLysGluPheLeuIleSerSerGlnAspGlyHisGlnTrpThrLeuPhePheGlnAs 1383
Qy 828 TGGTGTGGAGCAAGATATTTCAAGGAACAAGATTTATCACCAGGATGTGCGGTAA 887
Db 1383 nGly-----LysValLysValPheGlnGlyAsnGlnAspSerPheThrProValValas 1401
Qy 888 TAACCTTTTGGCCACCAATTTGACGTTTATTAAGTGAATCCTACCAATGCGACGA 947
Db 1401 nSerLeuAspProLeuLeuThrArgTyriLeuArgIleHisProGlnSerTrpValHi 1421
Qy 948 GAAATTCGCCATGAATAGGAGCTGCTCGGATGTCAG 984
Db 1421 sGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 1433
US-10-047-257-1
RESULT 8
US-10-047-257-1
```

Sequence 1, Application US/10047257
Patent No. US20020115152A1
GENERAL INFORMATION:
APPLICANT: Cho, Myung-Sam
APPLICANT: Chan, Sham-Tuen
APPLICANT: Kelsey, William
APPLICANT: Yee, Helena
TITLE OF INVENTION: Expression System for Factor VIII
FILE REFERENCE: MSB-7255.2
CURRENT APPLICATION NUMBER: US/10/047,257
CURRENT FILING DATE: 2002-01-15
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1438
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Derived from
OTHER INFORMATION: human factor VIII sequence
US-10-047-257-1

Alignment Scores:
Pred. No.: 2,22e-19 Length: 1438
Score: 316.00 Matches: 97
Percent Similarity: 44.44% Conservative: 51
Best Local Similarity: 29.13% Mismatches: 116
Query Match: 8.88% Indels: 69
DB: 12 Gaps: 12

US-10-060-830-1113 (1-1962) x US-10-047-257-1 (1-1438)

Qy 48 GGAGCTGGAGCCCA---GCAGCAATAGTCTGGTCTGGGTGCAAT----- 95
Db 1150 GlyGlnTrpAlaProLysLeuAlaArgLeuHisThrSerGlySerIleAsnAlaTrpSer 1169
Qy 96 -----GAACATTCATTAATCAAAAGCAATGAATCAATGCTGCTTCATGAGTGG 149
Db 1170 ThrLysGluProPheSerTrpIle-LysValAspLeuAlaProMetIleIleHisG1 1189
Qy 150 AATCCATGTTCTCGA-----CGCGGATTTTGGC 179
Db 1189 yileLysThrGlnGlyAlaArgGlnLysPheSerSerLeuTyriLeSerGlnPheIleI1 1209
Qy 180 CTCATATCTCTTATAGTAAACAGATCTAATTAATCTGTTGTCGACACTGCATCCATTT 239
Db 1209 eMetTySerLeuAspGlyLys----- 1216
Qy 240 TTTGGAACCTGAGTTCAGTAAGTAC-----TGCCAGCTGGTGTCTGCTTCCTTTTGC 293
Db 1217 -----LysTrpGlnThrTyArgGlyAsnSerThrGlyThrLeuMetValPhe-- 1232
Qy 294 TGATATCTGGAACAATTCCTCATGGATATAGAGATTCCTCGCCATTTGTCGATGGCTGG 353
Db 1233 -----PheGlyAsnValAspSerSer----- 1239
Qy 354 TGTGCATGCAGGAGTAGTGTCAACACACCTTGGCGGCGCAATCAGTGTGTGAATAGTAA 413
Db 1240 -----GlyIleLysHisAsnIlePheAsnProPheIle-----IleAlaAr 1253
Qy 414 AGGTATTCCTTATTAAGAACTCTTTTGGCTTAACAACGTCACATCTGTGGTGGGACACTT 473
Db 1253 gTyriLeArgLeuHisProThrHisTy-----SerIleArgSerThrLe 1268
Qy 474 ATCTACAGTCTTTTACATTTAACAAGTGAAGTGTATGAACACACTGGGAGTGGAGTC 533
Db 1268 uArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMetProLeuGlyMetGluSe 1288
Qy 534 TGGTGTGATCCGGATCCTCAATTAACAGCATCATCTGCTGCTGGAGTGGACACAC 593
Db 1288 rLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyRva-----Th 1303
Qy 594 AGGCAAGAGAACAGTGTGAACCCAAAGCCAGGCTGAAAAAACCCTGGA-----CC 647
Db 1303 rAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHisLeuGlnGlyArgSerAs 1323
Qy 648 GCCTTGGCTGCTTTTGGCCACTGATGATACCACTGGTTCATTAAGATTTGAATAAGGA 707
Db 1323 nAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGlnValAspPheGlnLysTh 1343
Qy 708 AAGAAATACAGGCATTAACCACTGGATCCACCATGCTGGAGGACACAAATTAATGT 767
Db 1343 rMetLysValThrGlyValThrThrGlnGlyValLysSerLeuLeuThrSerMetTyRva 1363
Qy 768 GTCTGCCTACAGATCCTGTACAGTGTAGTGGGACAGAAATGGACTGTGTACAGAGACC 827
Db 1363 lLysGluPheLeuIleSerSerGlnAspGlyHisGlnTrpThrLeuPhePheGlnAs 1383
Qy 828 TGGTGTGGAGCAAGATATTTCAAGGAACAAGATTTATCACCAGGATGTGCGGTAA 887
Db 1383 nGly-----LysValLysValPheGlnGlyAsnGlnAspSerPheThrProValValas 1401
Qy 888 TAACCTTTTGGCCACCAATTTGACGTTTATTAAGTGAATCCTACCAATGCGACGA 947
Db 1401 nSerLeuAspProLeuLeuThrArgTyriLeuArgIleHisProGlnSerTrpValHi 1421
Qy 948 GAAATTCGCCATGAATAGGAGCTGCTCGGATGTCAG 984
Db 1421 sGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 1433


```
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-957-641-2

Alignment Scores:
Pred. No.: 2,66e-19 Length: 2332
Score: 316.00 Matches: 97
Percent Similarity: 44.44% Conservative: 51
Best Local Similarity: 29.13% Mismatches: 116
Query Match: 8.88% Indels: 69
DB: 9 Gaps: 12

US-10-060-830-1113 (1-1962) x US-09-957-641-2 (1-2332)
QY 48 GGACGCTGAGCCCA---GCAGGCAATACTGTGGTCTGGGTGCAAAAT----- 95
Db 2044 GlyGlnTrpAlaProLysLeuAlaArgLeuHisTyrSerGlySerIleAsnAlaTrpSer 2063
QY 96 -----GAACCATTCATTAATCAAAAGGCAATGAATCATCTGCTTCATGAGTGG 149
Db 2064 ThrLysGluProPheSerTrpIle-LysValAspLeuLeuAlaProMetIleHisGlu 2083
QY 150 AATCATCTTCTCTGCA-----CCCGGATTTTGGC 179
Db 2083 yIleYsthrGlnGlyAlaArgGlnLysPheSerSerLeuTyrIleSerGlnPheIleI 2103
QY 180 CTCATACCTCTGTTATAGATAAACAAGATCAATTAATTAATTTGTTGGACACTGCATCAATTT 239
Db 2103 eMetTyrSerLeuAspGlyLys----- 2110
QY 240 TTTGGAACCTGAGTTCAGTAGTAC-----TGCCAGCTGGTGTCTGCTTCCTTTTGC 293
Db 2111 -----LysTrpGlnThrTyrArgGlyAsnSerThrGlyThrLeuMetValPhe-- 2126
QY 294 TGAGATATCTGGAACAATTCCTCATGATATAGAGATTCCTCGCCATTGTGCATGGCTGG 353
Db 2127 -----PheGlyAsnValAspSerSer----- 2133
QY 354 TGTGATGACGAGTAGTGTCAACACAGCTGGCGGCAATCAGTGTGTAATTAAGTAA 413
Db 2134 -----GlyIleLysHisAsnIlePheAsnProPheIle-----IleAlaAr 2147
QY 414 AGGTATTCCTTATATGAAAGTCTTTGGCTAACAAGCTCATCTGTGGTGGGACACTT 473
Db 2147 gTyrIleArgLeuHisProThrHisTyr-----SerIleArgSerThrLe 2162
QY 474 ATCTACAAGTCTTTTACATTAAGACAAGTGGATGTATGGAACACTGGGATGGAGTC 533
Db 2162 uArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMetProLeuGlyMetGluSe 2182
QY 534 TGGTGTATCCGGGATCCTCAATAACAGCATCATCTGTGCTGGAGTGGAGTACACAC 593
Db 2182 rLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyrPhe-----Th 2197
QY 594 AGGCAAGACAGACTGGAAACCCAAAGACCCAGCTGAAACACCTGCA-----CC 647
Db 2197 rAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHisLeuGlnGlyArgSerAs 2217
QY 648 GCCTTGGCTGCTTTGGCCACTGATACACAGTGGTTACAAATGAATTTGAATGAAGCA 707
Db 2217 nAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGlnValAspPheGlnLysTh 2237
QY 708 AAAGAAAATAACAGCATTAATACACTGGATCCACCATGGTGGAGCACAATTAATGT 767
Db 2237 rMetLysValThrGlyValThrThrGlnGlyValLysSerLeuLeuThrSerMetTyrVa 2257
```

```
QY 768 GTCTGCTCACAGAATCCTGTACAGTATGATGGCAGAAATGGACTGTGTACAGAGAGCC 827
Db 2257 lLysGluPheLeuIleSerSerSerGlnAspGlyHisGlnTrpThrLeuPhePheGlnAs 2277
QY 828 TGGTGTGGAGCAAGATAAGATATTTCAGAAACAAGATATATACAGGATGCGCTAA 887
Db 2277 nGly-----LysValLysValPheGlnGlyAsnGlnAspSerPheThrProValValAs 2295
QY 888 TAACCTTTTGGCCACCAATTAATTCACAGTGTATATAGTAGTGAATCCTACCAATGGCAGCA 947
Db 2295 nSerLeuAspProLeuLeuThrArgTyrLeuArgIleHisProGlnSerTrpValHI 2315
QY 948 GAAATTCATGCAATGAAATGGAGCTGCTCGGATGTGCAG 984
Db 2315 sGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 2327

RESULT 11
US-10-104-440-4
; Sequence 4, Application US/10104440
; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Qun
; TITLE OF INVENTION: ANTAGONISTS OF NEUROPILIN RECEPTOR FUNCTION AND USE
; FILE REFERENCE: 48802 C
; CURRENT APPLICATION NUMBER: US/10/104,440
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/580,803
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/069,155
; PRIOR FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 60/069,687
; PRIOR FILING DATE: 1997-12-29
; PRIOR APPLICATION NUMBER: 60/078,541
; PRIOR FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 931
; TYPE: PRT
; ORGANISM: human.
US-10-104-440-4

Alignment Scores:
Pred. No.: 7e-18 Length: 931
Score: 298.00 Matches: 171
Percent Similarity: 34.93% Conservative: 92
Best Local Similarity: 22.71% Mismatches: 291
Query Match: 8.37% Indels: 199
DB: 12 Gaps: 32

US-10-060-830-1113 (1-1962) x US-10-104-440-4 (1-931)
QY 67 GGCAAAATACGTGCTCTGGGTTCCAAATGAAC-----CATTCATTAATGAATCAAA 117
Db 102 GlyLysHisCysGly-----AsnIleAlaProProThrIleIleSerSer 116
QY 118 GGCAATGAAATACATTCGTCTTCATGAGTGAATCCATGTTCTGGACGCGGATTTTG 177
Db 117 GlySerMetLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGlyPheSer 136
QY 178 GCCTCATACTCTGTTATAGATAAACAAGATCAATTAATTAATTTGGACACTGCATCCAA 237
Db 137 LeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSerLysasn 152
QY 238 TTTTGGAACTGAGTTCAGTAAGTACTGCCACGCTGGTCTCTCTCTCTCTCTCTCTCTG 297
Db 153 PheThrSerProAsnGlyThrIleGluSerProGly-----PheProGlu 167
QY 298 ATATCTGGAACAATTCCTCATGGATATAGAGATTCCTCGCCATTTGTGCATGGCTGGTGTG 357
```

Db 168 -----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAlaLysPro 183
QY 358 CATGCAGGAGTAGTGTCAAAACAGTGTGGCGGCCAAATCAGT----- 399
Db 184 LysMetGluIleIleLeuGlnPheLeuIlePheAspLeuGluHisAspProLeuGlnVal 203
QY 400 -----GTYTAATTAGTAAGGATTCC----- 423
Db 204 GlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHisValGly 223
QY 424 -----TATTATGAAGTCTTTGGCTAACAAGCTCACATCTGTGGTGGCA 468
Db 224 ProLeuIleGlyLysTyrCysGlyThrLysThrProSerGluLeuArgSerThrGly 243
QY 469 CACTTATCTACAGTCTTTTACA----- 492
Db 244 IleLeuSerLeuThrPheHisThrAspMetAlaValAlaLysAspGlyPheSerAlaArg 263
QY 493 -----TTTAAGACAAGTGGATGTATGGAACACTGGGGATG 528
Db 264 TyrTyrLeuValHisGlnGluProLeuGluAsnPheGlnCysAsnValProLeuGlyMet 283
QY 529 GAGTCGTGTGTATCCGGATCCTCAATAACAGCATCATCTGTCTGTGGAGTGGACTGAC 588
Db 284 GluSerGlyArgIleAlaAsnGluGlnIleSerAlaSer----- 297
QY 589 CACACAGGCAAGAGAACAGTGTGGAACCCCAAAAGCCAGGCTGAAACACCTGGACCG 648
Db 298 ---ThrTyrSerAspGlyArgTrpTrpProGlnGlnSerArgLeuHisGlyAspAspAsn 316
QY 649 CTTGGCGCTTTCCTCCACTGATGATACCAGTGTGTACAAATAGATTGTAATAGGAA 708
Db 317 GlyTrpThrProAsnLeuAspSerAsnLysGluTyrLeuGlnValAspLeuArgPheLeu 336
QY 709 AAGAAATAACAGGCATTAACACTGGA-----TCCACCATGCTGGAGACAAATTAC 762
Db 337 ThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGluThrGlnAsnGlyTyr 356
QY 763 TATGTCTGCTCCATCAGATCTCTGACAGTGTATGGGACGAAATGAGTGTGTACAGA 822
Db 357 TyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAspTrpMetValTyrArg 376
QY 823 GAGCCTGGTGTGGAGCAAGATAGATATTCAAGGAACAAAGATATCACAGGATGTG 882
Db 377 HisGly-----LysAsnHisLysValPheGlnAlaAsnAlaThrGluValVal 394
QY 883 CGTAATAACTTTTGGCCACCAATTATGACGTTTATTAGAGTGAATCCTACCCATGG 942
Db 395 LeuAsnLysLeuHisAlaProLeuLeuThrArgPheValArgIleArgProGlnThrTrp 414
QY 943 CAGCAGAAATTCCTGAAATGGAGCTGCTCGGATGTCAGTTTATTCCTAAAGGTGCT 1002
Db 415 HisSerGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1003 CTCTCAAACTTACTCAACCTCCCTCTCGAAGCAGCAATGACCTCAAAACACTACA 1062
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMetLeuSerGlyLeuIle 445
QY 1063 GCCCTCCAAAATAGCCAAAGCTGTGCCCAAAATTT-----AGCAA 1107
Db 446 AlaAspSerGlnIleSerAlaSerThrGlnGluTyrLeuTrpSerProSerAlaAla 465
QY 1108 CCACCTACACCTCGCAGTCAATGATTTCTGCACAGACAGACAAACAACTGCCAGT 1167
Db 466 ArgLeuValSerSerArgSerGlyTrpPheProArgIleProGlnAlaGlnProGlyGlu 485
QY 1168 CCGATATCAGAAATACTACCGTAACCTCCAAATGTAACCAAGATGTA----- 1215
Db 486 GluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGlyValIleIleGlnGly 505
QY 1216 -----GGCTGGCTGCAGTTCCTTGTCCCTGTGTGTCATGGTCCCTCACT 1360
Db 506 AlaArgGlyGlyAspSerIleThrAlaValGluAlaArgAlaPheValArgLysPheLys 525

QY 1261 ACTCTCATTTCTCATATTAAGTGTGTCTGGCAGCTGG-----AGAAACAGA 1305
Db 526 ValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAspProArgThrGlnGln 545
QY 1306 AAGAAAAAATCGAAGGCACC-----TATGACTTACTTACTGGACCGCGCAGGTGG 1359
Db 546 ProLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1360 TGGAAAGGAATGAAGCAGTTTCTCTGCAAAAGCAGTGGACCATCAGGAAACCCAGTT 1419
Db 559 -----AspIleArgArgPheAspPro-----IleProAla 568
QY 1420 CCGTATAGCAGCAGCGAAGTTAATCACCTGAGTCCAGAGAGAGTCCACACATGCTGCAG 1479
Db 569 GlnTyrValArgValTyrProGluArgTrpSerProAlaGlyIleGlyMetArgLeuGlu 588
QY 1480 GCTGACTCTGCAGAGTAT-----GCTCAGCCACTGGTAGGAGGAATTGTTGTACACTT 1533
Db 589 ValLeuGlyCysAspTrpThrAspSerLysProThrValGluThrLeuGlyProThrVal 608
QY 1534 CATCAAGATCTACCTTTAAAGCA-----GAAGAAGGAAAGAACGAGC----- 1578
Db 609 LysSerGluGluThrThrProTyrProThrGluGluAlaThrGluCysGlyGlu 628
QY 1579 -----TATGCAGACTAGATCCTTACACTCACCAGGCGAGGAG----- 1618
Db 629 AsnCysSerPheGluAspLysAsp-LeuGlnLeuProSerGlyPheAsnCysAsnPh 648
QY 1619 -----TTTATCATGCTCT-----ATGCTGAACCACTCCCAA---TTAGGGGCGCT 1659
Db 648 eAspPheLeuGluGluProCysGlyTrpMetTyrAspHisAlaLysTrpLeuArg 666
QY 1660 GAGTATGCAACCCCAATCATCATGAGCATGTCTCAGGCGACCCACAA----- 1705
Db 667 -----ThrThrTrpAlaSerSerSerProAsnAspArgThrPhePr 681
QY 1706 -----CTTCA 1710
Db 681 OAspAspArgAsnPheLeuArgLeuGlnSerAspSerGlnArgGluGlyGlnTyrAlaAr 701
QY 1711 GTTGGTCAGCCCTCCACATCCACTTTCAGGCTACGGGACCAACCTCCCCACTAGTG 1770
Db 701 GluIleSerProPro-ValHisLeu-----ProArgSerProValCysM 716
QY 1771 GGAATTAACAATACATCTCTCTC----- 1793
Db 716 etGluPheGlnTyrGlnAlaThrGlyArgGlyValAlaLeuGlnValValArgGluA 736
QY 1794 -----CAGGACTGACAGCTGCTC-----CTCAGCCCGCCAGCCAGTATGATACCCGAA 1842
Db 736 laSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlyGlyGluTrpLys 755
QY 1843 GCTGGGAAGCCAGGTCTACCTGCCAGCAGGAATGGTGTACAGGTGCCACAGACACA 1902
Db 756 HisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGlnIle-----ValPhe 772
QY 1903 CAGAGAGTATCAGGACGAGGAGGATGGGAA 1935
Db 773 GluGlyValIleGlyLysGlyArgSerGlyGlu 783
RESULT 12
US-10-104-440-2
; Sequence 2, Application US/10104440
; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Qun
; TITLE OF INVENTION: ANTAGONISTS OF NEUROPILIN RECEPTOR FUNCTION AND USE
; FILE REFERENCE: THEREOF
; FILE REFERENCE: 48802 C
; CURRENT APPLICATION NUMBER: US/10/104,440

QY 1551 T---AAACCAAGAA 1563
Db 742 rGlnLysProGluGlu 747

RESULT 13
US-09-992-598-62
; Sequence 62, Application US/09992598
; Patent No. US20020160384A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paonl, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730P1C20
; CURRENT APPLICATION NUMBER: US/09/992,598
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252

; PRIOR FILING DATE: 1998-06-22
 ; PRIOR APPLICATION NUMBER: 60/090254
 ; PRIOR FILING DATE: 1998-06-22
 ; PRIOR APPLICATION NUMBER: 60/090349
 ; PRIOR FILING DATE: 1998-06-23
 ; PRIOR APPLICATION NUMBER: 60/090355
 ; PRIOR FILING DATE: 1998-06-23
 ; PRIOR APPLICATION NUMBER: 60/090429
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090431
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090435
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090444
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090445
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090472
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090535
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090540
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090542
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090557
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090676
 ; PRIOR FILING DATE: 1998-06-25
 ; PRIOR APPLICATION NUMBER: 60/090678
 ; PRIOR FILING DATE: 1998-06-25
 ; PRIOR APPLICATION NUMBER: 60/090690
 ; PRIOR FILING DATE: 1998-06-25
 ; PRIOR APPLICATION NUMBER: 60/090694
 ; PRIOR FILING DATE: 1998-06-25
 ; PRIOR APPLICATION NUMBER: 60/090695
 ; PRIOR FILING DATE: 1998-06-25
 ; PRIOR APPLICATION NUMBER: 60/090696
 ; PRIOR FILING DATE: 1998-06-25
 ; PRIOR APPLICATION NUMBER: 60/090862
 ; PRIOR FILING DATE: 1998-06-26
 ; PRIOR APPLICATION NUMBER: 60/090863
 ; PRIOR FILING DATE: 1998-06-26
 ; PRIOR APPLICATION NUMBER: 60/091360
 ; PRIOR FILING DATE: 1998-07-01
 ; PRIOR APPLICATION NUMBER: 60/091478
 ; PRIOR FILING DATE: 1998-07-02
 ; PRIOR APPLICATION NUMBER: 60/091544
 ; PRIOR FILING DATE: 1998-07-01
 ; PRIOR APPLICATION NUMBER: 60/091519
 ; PRIOR FILING DATE: 1998-07-02
 ; PRIOR APPLICATION NUMBER: 60/091626
 ; PRIOR FILING DATE: 1998-07-02
 ; PRIOR APPLICATION NUMBER: 60/091633
 ; PRIOR FILING DATE: 1998-07-02
 ; PRIOR APPLICATION NUMBER: 60/091978
 ; PRIOR FILING DATE: 1998-07-07
 ; PRIOR APPLICATION NUMBER: 60/091982
 ; PRIOR FILING DATE: 1998-07-07
 ; PRIOR APPLICATION NUMBER: 60/092182
 ; PRIOR FILING DATE: 1998-07-09

Alignment Scores:
 Pred. No.: 2,45e-11
 Score: 222.50
 Percent Similarity: 46.19%
 Best Local Similarity: 27.97%
 Query Match: 6.25%
 DB: 9
 Gaps: 10

US-10-060-830-1113 (1-1962) x US-09-992-598-62 (1-756)

QY 520 CTGGGATGGAGCTGGTGGTATCGCGGATCCTCAATAACAGCATATCTGTCTGTGAG 579

Db 139 LeuGlyLeuGluThrThrLeuLysIleThrAspPheGlnHisAlaSerThrValLysArg 158
 QY 580 TGG-----ACTGACACACAGGAGAAAGAACAGAGTGTGGAAACCCAAAGAGCCAGGCTG 633
 Db 159 TyrGlyLeuGlyAlaHisArgGlyArgLeuAsnIle-----GlnAlaGlyIle 174
 QY 634 AAAAAACCT-----GGACCCCTTGGCTCTTTTGGCCACTGATGATATACAGTGG 684
 Db 175 AsnGluAsnAspPheTyrAspGlyAlaIleTyrCysAlaGlyArgAsnAspLeuGlnIleTyr 194
 QY 685 TTACAAATAGATTGTAATAAGAAAGAAAGAACAGGCAATTAACACCTGGATCCACC 744
 Db 195 IleGluValAspAlaArgLeuThrArgPheThrGlyValIleThrGlnGlyArgAsn 214
 QY 745 ATGGTGGAGCACAATTACTATGTCTGCTCCCTACAGATCTCTGTACATGATGATGGCAG 804
 Db 215 SerLeuTyrLeuSerAspTyrValThrSerTyrLysValMetValSerAsnAspSerHis 234
 QY 805 AATGG---ACTGTGTACAGAGCCCTGGTGTGGAGCAAGATACATATTTTCAAGGAAC 861
 Db 235 ThrTyrValThrValLysAsnGlySerGly-----AspMetIlePheGluGlyAsn 251
 QY 862 AAGATTATCACCAGGATGTGCGTAATAACTTTTGGCCACCAATTTATTCACGTTTATT 921
 Db 252 SerGluLysGluIleProValLeuAsnGluLeuProValProMetValAlaArgTyrIle 271
 QY 922 AGAGTGAATCCTACCCCAATGG-----CAGCAGAAAAATTGCCATGAAAATGAGCTGCTC 975
 Db 272 ArgIleAsnProGlnSerTyrPheAspAsnGlySerIleCysMetArgMetGluIleLeu 291
 QY 976 GGATGTCAGTTTATTCCTAAAGTGTCTCCCTCCAAACCTTACTCAACCTCCACCTCTCGG 1035
 Db 292 GlyCysProLeu-----ProAspProAsn 299
 QY 1036 AAC-----AGCAATGACCTCAAAAACACTACAGCCCTCCCAAAATACCCAAA 1083
 Db 300 AsnTyrTyrHisArgArgAsnGluMetThrThrThrAspAspLeuAsp----- 315
 QY 1084 GTGTCGCCCAAAATTTACGCAACCACTACAACTCCAGTAGCAATGAAATTTCTCTGCA 1143
 Db 316 -----PheLysHisHisAsn-----Ty 321
 QY 1144 CAGACAGACAAACAACTGCCAGTCCCTCATATATCAGAAATACTACCG 1189
 Db 321 rLysGluMetArgGlnLeuMetLysValValAsnGluMetCysPro 336
 RESULT 14
 US-09-989-293A-62
 ; Sequence 62, Application US/09989293A
 ; Patent No. US20020177164A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ashkenazi, Avi J.
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, J. Christopher
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Kljavin, Ivar J.
 ; APPLICANT: Napier, Mary A.
 ; APPLICANT: Pan, James
 ; APPLICANT: Paoni, Nicholas F.
 ; APPLICANT: Roy, Margaret Ann
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K.

US-10-060-830-1113 (1-1962) x US-09-989-293A-62 (1-756)

```

QY 976 GGATGTCAGTTTATTCTCTAAAGTCCTCCCAAAACTTACTCAACCTCCACCTCCTCGG 103
Db 292 GlycysProLeu-----ProAspProAsn 299
QY 1036 AAC-----AGCAATGACCTCAAAAACACTACAGCCCTCCAAAATAATGCCAAA 1033
Db 300 AsnTyrTyrHisArgArgAsnGluMetThrThrThr-AspAspLeuAsp----- 315
QY 1084 GGTGCTGCCCCCAAAATTTACGCAACCACTACCAACACTCGCAGTAGCAATTTCTGTGCA 1143
Db 316 -----PheLysHisAsn-----Ty 321
RESULT 15
US-09-989-735-62
; Sequence 62, Application US/09989735
; Publication No. US20020193299A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730PIC61
; CURRENT APPLICATION NUMBER: US/09/989,735
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609

```

1	PRIOR APPLICATION NUMBER: 60/089900
2	PRIOR FILING DATE: 1998-06-18
3	PRIOR APPLICATION NUMBER: 60/089947
4	PRIOR FILING DATE: 1998-06-19
5	PRIOR APPLICATION NUMBER: 60/089948
6	PRIOR FILING DATE: 1998-06-19
7	PRIOR APPLICATION NUMBER: 60/089952
8	PRIOR FILING DATE: 1998-06-19
9	PRIOR APPLICATION NUMBER: 60/090246
10	PRIOR FILING DATE: 1998-06-22
11	PRIOR APPLICATION NUMBER: 60/090252
12	PRIOR FILING DATE: 1998-06-22
13	PRIOR APPLICATION NUMBER: 60/090254
14	PRIOR FILING DATE: 1998-06-22
15	PRIOR APPLICATION NUMBER: 60/090349
16	PRIOR FILING DATE: 1998-06-23
17	PRIOR APPLICATION NUMBER: 60/090355
18	PRIOR FILING DATE: 1998-06-23
19	PRIOR APPLICATION NUMBER: 60/090429
20	PRIOR FILING DATE: 1998-06-24
21	PRIOR APPLICATION NUMBER: 60/090431
22	PRIOR FILING DATE: 1998-06-24
23	PRIOR APPLICATION NUMBER: 60/090435
24	PRIOR FILING DATE: 1998-06-24
25	PRIOR APPLICATION NUMBER: 60/090444
26	PRIOR FILING DATE: 1998-06-24
27	PRIOR APPLICATION NUMBER: 60/090445
28	PRIOR FILING DATE: 1998-06-24
29	PRIOR APPLICATION NUMBER: 60/090472
30	PRIOR FILING DATE: 1998-06-24
31	PRIOR APPLICATION NUMBER: 60/090535
32	PRIOR FILING DATE: 1998-06-24
33	PRIOR APPLICATION NUMBER: 60/090540
34	PRIOR FILING DATE: 1998-06-24
35	PRIOR APPLICATION NUMBER: 60/090542
36	PRIOR FILING DATE: 1998-06-24
37	PRIOR APPLICATION NUMBER: 60/090557
38	PRIOR FILING DATE: 1998-06-24
39	PRIOR APPLICATION NUMBER: 60/090676
40	PRIOR FILING DATE: 1998-06-25
41	PRIOR APPLICATION NUMBER: 60/090678
42	PRIOR FILING DATE: 1998-06-25
43	PRIOR APPLICATION NUMBER: 60/090690
44	PRIOR FILING DATE: 1998-06-25
45	PRIOR APPLICATION NUMBER: 60/090694
46	PRIOR FILING DATE: 1998-06-25
47	PRIOR APPLICATION NUMBER: 60/090695
48	PRIOR FILING DATE: 1998-06-25
49	PRIOR APPLICATION NUMBER: 60/090696
50	PRIOR FILING DATE: 1998-06-25
51	PRIOR APPLICATION NUMBER: 60/090862
52	PRIOR FILING DATE: 1998-06-26
53	PRIOR APPLICATION NUMBER: 60/090863
54	PRIOR FILING DATE: 1998-06-26
55	PRIOR APPLICATION NUMBER: 60/091360
56	PRIOR FILING DATE: 1998-07-01
57	PRIOR APPLICATION NUMBER: 60/091478
58	PRIOR FILING DATE: 1998-07-02
59	PRIOR APPLICATION NUMBER: 60/091544
60	PRIOR FILING DATE: 1998-07-01
61	PRIOR APPLICATION NUMBER: 60/091519
62	PRIOR FILING DATE: 1998-07-02
63	PRIOR APPLICATION NUMBER: 60/091626
64	PRIOR FILING DATE: 1998-07-02
65	PRIOR APPLICATION NUMBER: 60/091633
66	PRIOR FILING DATE: 1998-07-02
67	PRIOR APPLICATION NUMBER: 60/091978
68	PRIOR FILING DATE: 1998-07-07
69	PRIOR APPLICATION NUMBER: 60/091982
70	PRIOR FILING DATE: 1998-07-07
71	PRIOR APPLICATION NUMBER: 60/092182
72	PRIOR FILING DATE: 1998-07-09

Alignment Scores:

Pred. No.: 2,45e-11 Length: 756
Score: 222.50 Matches: 66
Percent Similarity: 46.19% Conservatives: 43
Best Local Similarity: 27.97% Mismatches: 77
Query Match: 6.25% Indels: 50
DB: 9 Gaps: 10

US-10-060-830-1113 (1-1962) x US-09-989-735-62 (1-756)

```
QY 520 CTGGGATGAGTCGTGGTGTGATCGCGGATCTCTCAATAACACACATCATCTGTCTGGAG 579
      |||||:|||||: ||| ||| |||||:|||||
Db 139 LeuGlyLeuGluThrLeuLysIleThrAspGlnLeuHisAlaSerThrValLysArg 158
QY 580 TGG-----ACTCACACACAGGCAAGACAGACAGTTGGAAACCCAAAAGCCAGGCTG 633
      :: ||| |||||: |||
Db 159 TyrGlyLeuGlyAlaHisArgGlyArgLeuAsnIle-----GlnAlaGlyIle 174
QY 634 AAAAAACCT-----GGACCGCCTTGGCTGCTTTTCCCACTGATGAATACCACTGG 684
      ::| ||| ||| ||| |||||
Db 175 AsnGluAsnAspPheTyrAspGlyAlaTrpCysAlaGlyArgAsnAspLeuGlnGlnTrp 194
QY 685 TTACAATAGATTGTAATAAGAAAGAAATAACAGGCATTATAACCACTGGATCCACC 744
      |||||:|||||: ||| |||||:||||| |||
Db 195 IleGluValAspAlaArgArgLeuThrArgPheThrGlyValIleThrGlnGlyArgAsn 214
QY 745 ATGGTGGACACAATTAATCTATGCTGCTCCTACAGAACTCTGTACACTGATGATGGCAG 804
      ::| |||||:|||||: |||||: |||||: |||||
Db 215 SerLeuTriLeuSerAspTrpValThrSerTyrLysValMetValSerAsnAspSerHis 234
QY 805 AAATGG---ACTGTGTACAGAGCCCTGGTGTGGAGCAAGATAAGATATTTCAAGGAAAC 861
      ||| ||||| ||| ||| |||||:|||||
Db 235 ThrTrpValThrValLysAsnGlySerGly-----AspMetIlePheGluGlyAsn 251
QY 862 AAAGATTATCACCAGGATGCGCTAATACTTTTGGCCCAATAATTATTCACGCTTTTATT 921
      ::| ||| ||| |||||:|||||: |||||: |||||
Db 252 SerGluLysGluIleProValLeuAsnGluLeuProValProMetValAlaArgTyrIle 271
QY 922 AGAGTGAATCCTACCCNATGG-----CAGCAGAAATTCGCATGAAATGGAGCTGCTC 975
      |||||:||||| ||| ||| |||||:|||||: |||||: |||||
Db 272 ArgIleAsnProGlnSerTrpPheAspAsnGlySerIleCysMetArgMetGluIleLeu 291
QY 976 GGATGTCAGTTTATTCCTAAAGGTGCTCCTCCAAACCTTACTCAACCTCCACCTCCTCGG 1035
      ||||| ||||| ||| |||
Db 292 GlyCysProLeu-----ProAspProAsn 299
QY 1036 AAC-----AGCAATGACCTCAAAACACTACAGCCCTCCAAAATAAGCCAAA 1083
      ||| |||||: ||| |||
Db 300 AsnTyrTyrHisArgArgAsnGluMetThrThr-AspAspLeuAsp----- 315
QY 1084 GGTGCTGCCCAAAATTTACGCAACCACTACAACCTCGCAGTAGCAATGATTTCTCTGCA 1143
      ::| ||||| |||||: |||
Db 316 -----PheLysHisHisAsn-----Ty 321
QY 1144 CAGACAGAAACAACACTGCGCAGTCCTGATATCAGAAATACCTACCG 1189
      ::|||: ||||| |||||: |||||: |||||
Db 321 rLysGluMetArgGlnLeuMetLysValValAsnGluMetCysPro 336
```

Search completed: January 21, 2003, 09:52:35
Job time : 29.8461 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 21, 2003, 10:00:21 ; Search time 83.8719 Seconds
(without alignments)
3875.409 Million cell updates/sec

Title: US-10-060-830-3

Perfect score: 3888

Sequence: 1 MPEFLLLLVLLLELDAGA.....TOEVSGAGRGEGCDVKEIL 729

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 393868 seqs, 222934149 residues

Total number of hits satisfying chosen parameters: 787736

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPIO.spool/US10060830/runat_16012003_091559_29042/app_query.fasta_1.1742
-DB=PublishedApplications_NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10060830 -CGN_1_1.86 -runat_16012003_091559_29042
-NCPU=6 -ICPU=3 -NO_XLPYX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOPEXT=6
-FGAPEXT=7 -YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PTCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PTCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	3888	100.0	5657	9 US-09-974-298-96	/Sequence 96, Appl
2	1056	27.2	3151	9 US-10-003-132-1	Sequence 1, Appl
c 3	897	23.1	640	10 US-09-879-536-480	Sequence 480, App
4	838	21.6	580	10 US-09-864-761-12400	Sequence 12400, A

5	828	21.3	467	10	US-09-864-761-30344	Sequence 30344, A
6	801	20.6	2145	9	US-10-003-132-12	Sequence 12, Appl
7	575.5	14.8	2836	9	US-10-003-132-3	Sequence 3, Appl
8	574.5	14.8	1871	10	US-09-823-038A-43	Sequence 43, Appl
9	465	12.0	2868	9	US-10-003-132-5	Sequence 5, Appl
10	428	11.0	568	10	US-09-864-761-12562	Sequence 12562, A
11	415.5	10.7	1509	9	US-10-003-132-13	Sequence 13, Appl
12	412	10.6	228	10	US-09-864-761-29163	Sequence 29163, A
13	411	10.6	3404	12	US-10-104-440-3	Sequence 3, Appl
14	400.5	10.3	5653	12	US-10-104-440-1	Sequence 1, Appl
15	355	9.1	4599	9	US-09-974-298-167	Sequence 167, App
16	353.5	9.1	6909	10	US-09-880-107-2275	Sequence 2275, Ap
17	353.5	9.1	6925	9	US-10-115-563-13	Sequence 13, Appl
18	353.5	9.1	6925	9	US-10-115-563-26	Sequence 26, Appl
19	346	8.9	1270	10	US-09-880-107-3020	Sequence 3020, Ap
20	328	8.4	7914	12	US-10-095-718-3	Sequence 3, Appl
21	324	8.3	1374	9	US-10-003-132-14	Sequence 14, Appl
22	322	8.3	4629	10	US-09-150-811-7	GENERAL INFORMA
23	322	8.3	9009	9	US-09-957-641-1	Sequence 1, Appl
24	317	8.2	970	10	US-09-215-450-14	Sequence 14, Appl
25	308.5	7.9	4999	10	US-10-007-968-14	Sequence 14, Appl
26	308.5	7.9	4999	10	US-09-740-211-14	Sequence 1, Appl
27	308.5	7.9	7944	12	US-10-095-718-1	Sequence 13, Appl
28	308.5	7.9	11933	9	US-10-007-968-13	Sequence 13, Appl
29	308.5	7.9	11933	10	US-09-740-211-13	Sequence 1, Appl
30	231	5.9	2490	10	US-09-900-518A-1	Sequence 61, Appl
31	220.5	5.7	3772	9	US-09-992-598-61	Sequence 61, Appl
32	220.5	5.7	3772	9	US-09-989-293A-61	Sequence 61, Appl
33	220.5	5.7	3772	9	US-09-989-735-61	Sequence 61, Appl
34	220.5	5.7	3772	9	US-09-990-444-61	Sequence 61, Appl
35	220.5	5.7	3772	9	US-09-989-730-61	Sequence 61, Appl
36	220.5	5.7	3772	9	US-09-990-436-61	Sequence 61, Appl
37	220.5	5.7	3772	9	US-09-991-181-61	Sequence 61, Appl
38	220.5	5.7	3772	9	US-09-993-687-61	Sequence 61, Appl
39	220.5	5.7	3772	9	US-09-989-734-61	Sequence 391, App
40	220.5	5.7	3772	9	US-10-028-072-391	Sequence 61, Appl
41	220.5	5.7	3772	10	US-09-997-653-61	Sequence 61, Appl
42	220.5	5.7	3772	10	US-09-989-722-61	Sequence 61, Appl
43	220.5	5.7	3772	10	US-09-989-723-61	Sequence 61, Appl
44	220.5	5.7	3772	10	US-09-989-279-61	Sequence 61, Appl
45	220.5	5.7	3772	10	US-09-989-727-61	Sequence 61, Appl

ALIGNMENTS

RESULT 1
US-09-974-298-96
; Sequence 96, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 96
; LENGTH: 5657
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 241227.17
; NAME/KEY: unsure
; LOCATION: 4516
; OTHER INFORMATION: a, t, c, g, or other
US-09-974-298-96

Alignment Scores:
Pred. No.: 0 Length: 5657


```
Db 1498 GGAATCTTTGACCTTT-----AGAAAGAAAGAAAGAAA 1533
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 517 GlyThrTyAspLeuProTyrTrp-----AspArgAlaGlyTrpTyrLysGly 532
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 1534 GGAAGT-----CCGATGATGACAGAGGCTCAGAAACACAGACTGTGGAGCAG 1584
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 533 MetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSer 552
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 1585 ATTAATATCCCTTT-----GCCAGACATCAG 1611
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 553 SerSerGlu-----ValAsnHisLeuSerProArgGluValThr-----ThrVal 567
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 1612 TCAGCTGAGTTTACCATCAGCTATGATGATGAGAGGAGATGACACAAAGTTAGATCTC 1671
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 568 LeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyLeuValGlyThrLeu 587
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 1672 ATCACAAGTGATGTCAGATATACAGCAGCCCTCATGATTGGACCCGGNACATCAG 1731
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 588 HisGlnArgSerPheLysPro-----GluGluGlyLysGluAlaGlyTyrAlaAsp 605
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 1732 AGGAAGGGCTCCACCTCCGGCCCATGGACACGGATGCCGAGGAGCAGGG---GTGAGC 1788
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 606 LeuAspPro-----TyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGlu 622
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 1789 ACCGATGCGGGCGGCACATGACTGCCCCGCGAGCGGGCCGCCGACGAGTACGCGCTG 1848
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 623 ProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetAspMetSerGlyHis 642
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 1849 CCCCTGGGCCCCCGAGCGGAGTACGCCAGCCATCGCTG-----GAGCGGCAC 1899
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 643 ProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnPro 662
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 1900 GTGCTGCGCGCCACACAGTCTCTCGCAGAGCGGCTACCGCTCCAGGCGCCCGAGCCC 1959
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 663 -----ProProLeu-----ValGlyThrTyr 669
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 1960 GGCACAAACACTCCCTCTCCGCGGGCGGCTTCTCCCGCTAGCGGGTGGCGGCCAG 2019
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 670 AsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrPro 689
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 2020 GACGAGACTCAAGGCCACACAGCCACAGCTCGCGGACAGGGGTACGCGGGCCC 2079
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 690 LysAla-GlyLysProGlyLeuProAlaProAspGluLeuValTyrGlnValProGln 708
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 2080 AAAGTGTCAGCGCCCTCGCCACCGCAAGCGGACACACCTGACTCTCAGAAAGCCCCCAA 2137
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480
```

RESULT 3

```
US-09-879-536-480/c
; Sequence 480, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/879,536
; PRIOR FILING DATE: 2001-09-21
; PRIOR FILING DATE: US 60/088,801
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
```

RESULT 4

```
US-09-864-761-12400
; Sequence 12400, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
```

```
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Alignment Scores:
Pred. No.: 4,99e-76 Length: 640
Score: 897.00 Matches: 194
Percent Similarity: 91.98% Conservative: 1
Best Local Similarity: 91.51% Mismatches: 17
Query Match: 23.07% Indels: 6
DB: 10 Gaps: 0

US-10-060-830-3 (1-729) x US-09-879-536-480 (1-640)

QY 375 LeuProIleIleAlaArgPheIleArgValAlaAsnProThrGlnTrpGlnGlnLysIle 394
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 630 TTGCCACCAATATATGGCAGC-TTATTAGAGT-AATCCTTACCCATGGCAGCAGAAAT- 574
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 395 AlaMetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeu 414
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 573 GCCATGAAAT-GAACTGCTCGGATGTCAGTTATTCTNTAAAGT-NGTCNTCCAAACTT 516
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 415 ThrGlnProProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLys 434
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 515 ACTCAACCTCCACCTCTTNGNAACAGCAATNGCCTCAAAACACACTACAGCCCTCCAAA 456
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 435 IleAlaLysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGlu 454
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 455 TA-GCCAAAGGTGTCGCCCCAAAATTTACGCAACCACTTACCACTTCGCAGTAGCAATGAA 397
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 455 PheProAlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThr 474
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 396 TTTCTGCGACAGACAGACAACTCCAGTCTGTATATCAGAAATACTACCGTAAC 337
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 475 ProAsnValThrLysAspValAlaLeuAlaValLeuValProValLeuValMetVal 494
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 336 CCAATGTATACAAAGATGTAGCGCTGGCTGCTTCTGTCCCTGCTGCTGCTGCTGCTGCT 277
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 495 LeuThrThrLeuIleLeuLeuValCysAlaThrHisTrpArgAsnArgLysLysLys 514
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 276 CTCCTACTCTCATCTCATATTAGTGTGCTTGGCTGGGACTGGAGAAACAGAAAGAAAA 217
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 515 ThrGluGlyThrThrAspLeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLys 534
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 216 ACTGAAGGCACCTATGACTTACTTACTGGGCGGGCAGCTTGTGGAAAGGAATGAAG 157
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 535 GlnPheLeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSerSerSer 554
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 156 CAGTTTCTTCTCGCAAAAGCAGTGGACCATGAGGAAACCCAGTTCGCTATAGCAGCAGC 97
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 555 GluValAsnHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGlu 574
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 96 GAAGTTATACCTGAGTCCCAAGAGAGTACACAGTGTGCTGAGGCTGCTGCTGCTGCTGCT 37
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

QY 575 TyrAlaGlnProLeuValGlyGlyLeuValGlyThr 586
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Db 36 TATGCTCAGCCACTGGTAGGAGGAATGTTGGTACC 1
```

```

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US 09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12400
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
; US-09-864-761-12400

Alignment Scores:
Pred. No.: 1,61e-70 Length: 580
Score: 838.00 Matches: 161
Percent Similarity: 95.38% Conservative: 4
Best local Similarity: 93.06% Mismatches: 7
Query Match: 21.55% Indels: 1
DB: 10 Gaps: 0

US-10-060-830-3 (1-729) x US-09-864-761-12400 (1-580)

Qy 558 HisLeuSerProArgGluValThrValLeuGlnAlaAsp-SerAlaGluTyrAlaG1 577
      ::::: ::::: ::::: :::::
Db 21 AATCTCTTACTAGTAAATGACTTTCTGTTGTTGATATCCACAGATGCTCA 80

Qy 577 nProLeuValGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluG1 597
      ::::: ::::: ::::: :::::
Db 81 GCACCTGGTAGGAGGAATTTGGTGTACATCTCAAGATCTACCTTTAAACCAAGA 140
```

```

Qy 597 uGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTy 617
      ::::: ::::: ::::: :::::
Db 141 AGGAAACAGCAGGAGCTATGCAGACCTAGATCCTTACAACTCACCAGGCGCAGGAAGTTTA 200

Qy 617 rHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMe 637
      ::::: ::::: ::::: :::::
Db 201 TCATGCCCTATGCTGAACCACTCCCAATTTACGGGGCTGAGTATGCAACCCCAATCATCAT 260

Qy 637 tAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAl 657
      ::::: ::::: ::::: :::::
Db 261 GGACATGTCAGGGCACCCCACTTCAGTTGGTCAGGCGCTCCACATCCACTTTCACAGGC 320

Qy 657 aThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAs 677
      ::::: ::::: ::::: :::::
Db 321 TACGGGGAACCAACCTCCCACTAGTGGAACTTACAAATACACTTCTCTCCAGGACTGA 380

Qy 677 pSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuPr 697
      ::::: ::::: ::::: :::::
Db 381 CAGCTGCTCCTCAGCCAGGCGCCAGTATGATACCCGGAAGCTGGGAAGCAGGCTCTACC 440

Qy 697 oAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaG1 717
      ::::: ::::: ::::: :::::
Db 441 TGCCCCAGACGAATTTGGTGTACCGAGTGCCACAGACACACAAGAAGTATCAGGAGCAGG 500

Qy 717 yArgAspGlyGluCysAspValPheLysGluIleLeu 729
      ::::: ::::: ::::: :::::
Db 501 AAGGATGGGAATGTGATGTTTTTAAAGAAATCCTT 537

RESULT 5
; Sequence 30344, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US 09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12400
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
; US-09-864-761-12400
```

FILE REFERENCE: 00-62
CURRENT APPLICATION NUMBER: US/10/003.132
CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/249,004
PRIOR FILING DATE: 2000-11-15
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 2145
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: degenerate nucleotide sequence
NAME/KEY: misc_feature
LOCATION: (1)...(2145)
OTHER INFORMATION: n = A,T,C or G
US-10-003-132-12

Alignment Scores:
Pred. No.: 3.5e-66 Length: 2145
Score: 801.00 Matches: 211
Percent Similarity: 41.12% Conservative: 83
Best Local Similarity: 29.51% Mismatches: 321
Query Match: 20.60% Indels: 100
DB: 9 Gaps: 21

US-10-060-830-3 (1-729) x US-10-003-132-12 (1-2145)

Qy 21 GlnGlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSer 40
Db 106 GARYTNGNGAYGGNTGYGNCAYTNGTACATAYCARGAYWSNGNACATGACNWSN 165
Qy 41 IleAsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluLeuArgValLys 60
Db 166 AARAYTAYCCNGGACNACATAYCCNAAAYCAYACNGTGTGARAARACNATHACNGTNCN 225
Qy 61 MetGlyGluAlaGlyValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCys 80
Db 226 AARGNNAARMGNMTNATHYTNMGNTNGNGAYTNGAYATHGAR---WSNACACNTGY 282
Qy 81 HisPheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluLeuLys 100
Db 283 GCNWSNGAY-----TAYTNTNTTYACNWSNWSNGAYCARTAYGNCN 330
Qy 101 TyrCysGlyLeuGlyLeuGluMetAsnHisSerIleGluSerLysGlyAsn-----Glu 118
Db 331 TAYTGIGN-----WSNATGACNCTNCCNAAARGARYTNTYNTNAYACNWSNGAR 381
Qy 119 IleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyr 138
Db 382 GTNACNGTNGMTTYGARNWSNGNWSNCAYATHWSNGNMGNGNTTYTNTNACNTAY 441
Qy 139 SerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGlu 158
Db 442 GCNWSNGAYCAYCCNGAYTNTACNTGTYTNGARMGNCNWSNCAYTAYTNAAR 501
Qy 159 ProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluLeuSerGly 178
Db 502 ACNGARTAYWSNAARTTYTGCCNGCNGNTGTMNGNGAYGTNGCNGNGAYATHWSNGN 561
Qy 179 ThrIleProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGly 198
Db 562 AAYATGGTNGAYGGNTAYMNGNAYACNWSNTYNTNTGYAARGCNCNATHACGCGNGN 621
Qy 199 ValValSerAsnThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyr 218
Db 622 ATHATGCGNGAYCARYTNGGNGCNCARATHWSNGTNTTCARMGNAARGNATHWSNMG 681
Qy 219 TyrGluSerSerLeuAlaAsnValThrSerValValGlyHisLeuSerThrSerLeu 238
Db 682 TAYGARGGNATHYTNCCNAAAYGGNTNTYTNWSNMGNGAYGGNWSNTYTNWSNGAYAA 741
Qy 239 PheThrPheLysThrSerGlyCystyrGlyThrLeuGlyMetGluSerGlyValIleAla 258

PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 30344
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC013497.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
OTHER INFORMATION: EST_HUMAN HIT: AW583777.1, EVALUE 3.00e-51
OTHER INFORMATION: NT HIT: Z70177.1, EVALUE 2.90e-01
OTHER INFORMATION: SWISSPROT HIT: Q9Y011, EVALUE 7.40e-01
US-09-864-761-30344

Alignment Scores:
Pred. No.: 1.03e-69 Length: 467
Score: 828.00 Matches: 155
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21.30% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-3 (1-729) x US-09-864-761-30344 (1-467)

Qy 575 TyrAlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLys 594
Db 3 TATGCTCAGCCACTGGTAGGAGGAATTGTGTACACTTCATCAAGATCTACCTTTAA 62
Qy 595 ProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGln 614
Db 63 CCAGAAGAAGAAAGAACGAGCTATGACACCTAGATCCTTACAACTCACCAGGCGAG 122
Qy 615 GluValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrPro 634
Db 123 GAAGTTATATGCTGCTATCTGAACCACTCCCAATACGGGCGCTGAGTATGCAACCCA 182
Qy 635 IleIleMetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThr 654
Db 183 ATCATCATGGACATGTCAGGGGACCCCACTTTCAGTTGTCAGCCCTCCACATCCACT 242
Qy 655 PheLysAlaThrGlyAsnGlnProProProLeuValGlyThrTyrAsnThrLeuLeuSer 674
Db 243 TTCAAGGCTACGGGGAACCAACCTCCCCACTAGTGGGAACCTACAATACACTCTCTCC 302
Qy 675 ArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysPro 694
Db 303 AGACTACGACAGCTGCTCCCTCAGCCAGGCCCACTGATGATACCCGAAAGCTGGAGCCA 362
Qy 695 GlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSer 714
Db 363 GGCTACCTGCCCCAGACGAGATGTTGTACAGGTGCCCCACAGGACACAGAAGTATCA 422
Qy 715 GlyAlaGlyArgAspGlyGluCysAspValPheLysGluIleLeu 729
Db 423 GGAGCAGGAAGGATGGGGAATGTGATGTTTTTAAAGAAATCCTT 467

RESULT 6
US-10-003-132-12
Sequence 12, Application US/10003132
Publication NO. US20020192750A1
GENERAL INFORMATION:
APPLICANT: Fox, Brian A.
APPLICANT: Gao, Zeren
APPLICANT: Shoemaker, Kimberly E.
TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5

Db 1666 WSNACNTYTMGNCNCATGGAYACNAYCGNARGARGCNGCN--GTVNSNACNGAYCCN 1772
 Qy 609 -----TyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuPro 625
 Db 1723 GNGGNCAYTAYGAYTCYCCNCAARMGNCNGNMGNCAYGARTAYGCGNYTNCNTNGCN 1782
 Qy 626 IleThrGlyProGluTyrAlaThrProIleIleMetAspMetSerGlyHisProThrThr 645
 Db 1783 CCNCCNGARCCNGARTAYGCNACNCCNAYTHGTN-----GARMGNCAYGTNTNTMGN 1833
 Qy 646 SerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnPro----- 662
 Db 1834 GCNCAYACNTYYWSNGCNCAWSNGNTAYMGNWTNCNGGCCNCCARCCNGNCAYAA 1899
 Qy 663 -----ProProLeu-----ValGlyThrTyrAsnThrLeu 672
 Db 1894 CAYWSNTNWSNWSNGGNGGNTYYWSNCCNGTNGCNGGNGTNGGNCNCARGAYGGNGAY 1955
 Qy 673 LeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAla--- 691
 Db 1954 TAYCARMGNCNCAYWSNGCNCAARMGNGNTAYMGNMGNTAYGAYMGNCCNAAACGNGTN 2013
 Qy 692 -----GlyLysProGlyLeuProAlaPro 699
 Db 2014 WSNCCNTYTCNACNARGWSNGNCAYCCNGAYWSNCAARAACCN 2058

RESULT 7
 US-10-003-132-3
 ; Sequence 3, Application US/10003132
 ; Publication No. US20020192750A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fox, Brian A.
 ; APPLICANT: Gao, Zeren
 ; APPLICANT: Shoemaker, Kimberly E.
 ; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
 ; FILE REFERENCE: 00-62
 ; CURRENT APPLICATION NUMBER: US/10/003,132
 ; CURRENT FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/249,004
 ; PRIOR FILING DATE: 2000-11-15
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 2836
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (129)...(1640)
 US-10-003-132-3

Alignment Scores:
 Pred No.: 1,05e-44 Length: 2836
 Score: 575.50 Matches: 183
 Percent Similarity: 35.74% Conservative: 84
 Best Local Similarity: 24.50% Mismatches: 181
 Query Match: 14.80% Indels: 299
 DB: Gaps: 20

US-10-060-830-3 (1-729) x US-10-003-132-3 (1-2836)
 Qy 2 ProLeuPheLeuLeuLeuLeuValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 21
 Db 150 CCGAGTGTCCTGGCGCTGCTGTGTCGCGCTGCTGTCCTCGGTGCGAGCGGAGGAG 209
 Qy 22 GlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrThrThrSerIle 41
 Db 210 CTGGGTGATGCTGTGGGCACATAGTACCTCTCAGGACAGTGGCACAATGCATCTAAG 269
 Qy 42 AsnTyrProGlnThrTyrProAsnSerThrValCysGlnTrpGluIleArgValLysMet 61
 Db 270 AATTATCCAGGAGCTTACCCCAATACACTGTGTGTGAAGAAGATCATCAGCTCCCAAG 329

QY 62 GlyGluArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCysHis 81
Db 330 GGAAGAGACTTATCTGAGGTGGAGATTTCAACATTGAG---TCCAAGACCTGCGCT 386
QY 82 PheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLysTyr 101
Db 387 TCTGACTATCTCTCTCAGCAGT-----GCAACAGATCAGTATGTCATAT 434
QY 102 CysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThrLeu 121
Db 435 TGTGGG---AGTTGGGCTGTCCTCAAGAACTCCGGCTCAACTCAACGAAGTACTGTC 491
QY 122 LeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValle 141
Db 492 CTCCTCAAGAGTGATCTCAGATTTCTGGCGGGGCTTCTGCTGACCTACCCAGCAGT 551
QY 142 AspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPhe 161
Db 552 GACCATCCAGATTATAACCTGTTTGGAAACGAGCGCCATATTTCAGGAAATAC 611
QY 162 SerLysTyrCysProIleGlyCysLeuLeuProPheAlaGluIleSerGlyThrIlePro 181
Db 612 AGCAATCTGCCAGCTGGCTAGACATAGCAGGAGATATTCTGGGAATACAAA 671
QY 182 HisGlyTyrArgAspSerProLeuCysMetAlaGlyValHisAlaGlyValValSer 201
Db 672 GATGGTTACAGAGATACCTCTTATTGTGCAAGCTGCATCCAGCGAGGATCATCA 731
QY 202 AsnThrLeuGlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGluSer 221
Db 732 GATGAATAGGTGGCCACATCAACTGCTTCAGACCAAGGATAGTCACTATGAAGA 791
QY 222 SerLeuAlaAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThrPhe 241
Db 792 CTCCTGGCCAAAGGGGTGCTCTCCGCGCATGTTCTTCTCGGAAAGCGATTTCTTTT 851
QY 242 LysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspProGln 261
Db 852 ACNACC----- 857
QY 262 IleThrAlaSerSerValLeuGluThrThrAspHisThrGlyGlnGluAsnSerTrpLys 281
Db 857 ----- 857
QY 282 ProLysLysAlaArgLeuLysLysProGlyProProTrpAlaAlaPheAlaThrAspGlu 301
Db 857 ----- 857
QY 302 TyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThrThr 321
Db 857 ----- 857
QY 322 GlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAsp 341
Db 857 ----- 857
QY 342 AspGlyGlnLysTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGln 361
Db 858 -----CCAGGAATG----- 866
QY 362 GlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProProIleIleAlaArg 381
Db 866 ----- 866
QY 382 PheIleArgValAsnProThrGlnTrpGlnLysIleAlaMetLysMetGluLeuLeu 401
Db 866 ----- 866
QY 402 GlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProProArg 421
Db 866 ----- 866
QY 422 AsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaPro 441

Db 866 ----- 866
QY 442 LysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGln 461
Db 866 ----- 866
QY 462 ThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAspVal 481
Db 867 -----AATATATAC-----ACTGTG 881
QY 482 AlaLeuAlaAlaValLeuValProValLeuValMetValLeuThrThrLeuLeuLeu 501
Db 882 GCGATTTCATCAGTGATC-----TTCAATCGCCCTCTCTGACTGGAATGGGATCTTT 935
QY 502 LeuValCysAlaTrpHisTrpArgAsnArgLysLysThrGluGlyThrTyrAspLeu 521
Db 936 GCATCTGT-----AGAAAGAGAAAAGAAAGGAAT----- 968
QY 522 ProTyr-----TrpAspArgAlaGlyTrpLysGlyMetLysGlnPheLeu 537
Db 969 CCATATGTCTCAGCTGACCTCAGAAAACAGGCTGTGGAGCAGATTAATATATCCTTT 1028
QY 538 ProAlaLysAlaValAspHisGluGluThrProValArgTyrSerSerGlu----- 555
Db 1029 -----GCCAGGCATCAGTCGACGGAATTTACC 1055
QY 556 ValAsnHisLeuSerProArgGluValThr-----ThrValLeuGlnAlaAspSer 572
Db 1056 ATCAGCTATGACATGAAAAGAGATGACACAAAAGTTGGATCTCATCTAGTATATG 1115
QY 573 AlaGluTyrAlaGlnProLeuValGlyIleValGlyThrLeuHisGlnArgSerThr 592
Db 1116 GCAGATTATCAGCAGCTCTCATGATTGGCAGAGCAGCTGCCGAGAAGGCTCTACC 1175
QY 593 PheLysPro-----GluGlu-----GlyLysGluAlaGlyTyr 603
Db 1176 TTCCGACCCCTGGACACAGACACTGAGGAGTGCAGAGTGAACACTGAGGCCAGGCCAC 1235
QY 604 AlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluPro 623
Db 1236 -----TATGACTGTCTCACCGCGCGGCGGCGCATGAGTACGACGCTGCCT 1280
QY 624 LeuProIleThrGlyProGluTyrAlaThrProIleIle----- 636
Db 1281 TTGAGCGCACTCAGAACCTGAGTATGCCACACTATGCTGGAGCGGCACCTGTCGAGCT 1340
QY 637 -----MetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSer 653
Db 1341 CACACCTTCTCCACACAGAGCGGCTACCGAGTCCCTGGG---CCCAGGCCACTCACAAA 1397
QY 654 ThrPheLysAlaThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeu 673
Db 1398 CACTCCCATTCCTCGGAGGCTTCTCTCTGCTACAGGAGCCACCCAGGTTGAAAGCTAT 1457
QY 674 SerArgThrAspSerCysSerSerAlaGlnAlaGlnThrAspThrProLysAlaGlyLys 693
Db 1458 CAGAGCCAGCAAGCCCAAGCCTGTGGTGGTGGCTATGACAAGCCT---GCTGCTAGC 1514
QY 694 ProGlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluVal 713
Db 1515 AGCTTCTTGGACAGCAGAGAC-----CCAGCCTCTCAGTCACAGATG 1556
QY 714 SerGlyAlaGlyArgAspGly 720
Db 1557 ACTTCCGGGGGAGATGATGGT 1577

RESULT 8

US-09-823-038A-43
; Sequence 43, Application US/09823038A
; Patent No. US2002005835A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna

```

; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevin
; APPLICANT: Onrust, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Murison, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.103763
; CURRENT APPLICATION NUMBER: US/09/823,038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 1871
; TYPE: DNA
; ORGANISM: Mouse
US-09-823-038A-43

Alignment Scores:
Pred. No.: 7,07e-45 Length: 1871
Score: 574.50 Matches: 183
Percent Similarity: 35.74% Conservatives: 84
Best Local Similarity: 24.50% Mismatches: 181
Query Match: 14.78% Indels: 299
DB: 20 Gaps: 20

US-10-060-830-3 (1-729) x US-09-823-038A-43 (1-1871)

Qy 2 ProLeuPheLeuLeuLeuLeuValLeuLeuLeuLeuLeuLeuGluAspAlaGlyAlaGln 21
Db 85 CCGAGTCTCTGGCGCTGCTGTCGCGGTGCTCGCTCCGCTTGCAGCGGAGGAG 144
Qy 22 GlnGlyAspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerIle 41
Db 145 CTGGGTGATGCTGGGCATAGTAGNCCTCTCAGGACAGTGGCACAATGACATCTAAG 204
Qy 42 AsnTyrProGlnThrTyrProAsnSerThrValCysGluTrpGluIleAArgValLysMet 61
Db 205 AATTATCCAGGAGCTTACCCCAATATACACTGTGTGTAAGATCATCAGCTCCCAAG 264
Qy 62 GlyGluArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCysHis 81
Db 265 GGGAGAGACATTATCTGAGCTGGGAGATTGAACTAGAG---TCCAAGACCTGCCCT 321
Qy 82 PheAsnTyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIleGlyLysTyr 101
Db 322 TCTGACTATCTCCTTCAGCAGT-----GCAACAGATCAGTATGGTCCATAT 369
Qy 102 CysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThrLeu 121
Db 370 TGTGGG---AGTTGGGCTGTTCCCAAGAACTCCGGCTGAACTCAAAAGAAAGTGACTGTC 426
Qy 122 LeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIle 141
Db 427 CTCTTCAAGAGTGGATCTCATTCTGCGCGGCTTCTGCTGACCTACGCCAGCAGT 486
Qy 142 AspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPhe 161
Db 487 GACCATCCAGATTAAATAACCTGTTTGGACAGCAGGCCCATATTATTCGAGGAAAAATAC 546
Qy 162 SerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIlePro 181
Db 547 AGCAAAATCTCCAGCTGGCTGTAGACATAGCAGCATATATTCTGGGAATACAAA 606
Qy 182 HisGlyTyrArgAspSerProLeuCysMetAlaGlyValHisAlaGlyValValSer 201
Db 607 GATGTTACAGATGATATCTTTATGTGCAAGCTGCCATCCAGCGAGGATCATCACA 666
Qy 202 AsnThrLeuGlyGlnIleSerValIleSerLysGlyIleProTyrTyrGluSer 221
Db 667 GATGAACATAGTGGCCACATCACTTGTCTCAGACGAAAGGATTAAGTCACTATCAAGA 726
Qy 222 SerLeuAlaAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThrPhe 241

```

```

Db 727 CTCTGGCCAAATGGCGTCTCTCCGGCATGGTTCTTGTGGAAAGCCGATTTCTTTT 786
Qy 242 LysThrSerGlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspProGln 261
Db 787 ACAACC----- 792
Qy 262 IleThrAlaSerSerValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrpLys 281
Db 792 ----- 792
Qy 282 ProLysLysAlaArgLeuLysLysProGlyProProTrpAlaAlaPheAlaThrAspGlu 301
Db 792 ----- 792
Qy 302 TyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThrThr 321
Db 792 ----- 792
Qy 322 GlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAsp 341
Db 792 ----- 792
Qy 342 AspGlyGlnLysTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGln 361
Db 793 -----CCAGGAATG----- 801
Qy 362 GlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProIleIleAlaArg 381
Db 801 ----- 801
Qy 382 PheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeu 401
Db 801 ----- 801
Qy 402 GlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProProArg 421
Db 801 ----- 801
Qy 422 AsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaPro 441
Db 801 ----- 801
Qy 442 LysPheThrGlnProLeuGlnProArgSerAsnGluPheProAlaGlnThrGluGln 461
Db 801 ----- 801
Qy 462 ThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAspVal 481
Db 802 -----AATATTACA---ACTGTG 816
Qy 482 AlaLeuAlaAlaValLeuValProValLeuValMetValLeuThrLeuLeuLeuIle 501
Db 817 GCGATTCATCACTGATC-----TTCATCGCCCTCTCTGACTGGAATGGGATCTTT 870
Qy 502 LeuValCysAlaThrHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrAspLeu 521
Db 871 GCAATCTGT-----AGAAAGAGGAAAAAGAGAGAAAT----- 903
Qy 522 ProTyr-----TrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeu 537
Db 904 CCATATGTGTACAGCTCAGCCTCAGAAACAGGCTTTGGAAGCAGCATTAATATATCCCTTT 963
Qy 538 ProAlaLysAlaValAspHisGluThrProValArgTyrSerSerSerGlu----- 555
Db 964 -----GCCAGGCATCCAGCGGAATTACC 990
Qy 556 ValAsnHisLeuSerProArgGluValThr-----ThrValLeuGlnAlaAspSer 572
Db 991 ATCAGCTATGACAATGAAAGAGATGACACAAAAAGTTGGATCTCATCTAGTGATG 1050
Qy 573 AlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThr 592

```


QY 181 ATGGGAGAGAGTTGCGCATCAAAATTTGGTGAATTTGACATTTGAAGATTTCTGATTTCTTGT 240
|||||
Db 181 ATGGGAGAGAGTTGCGCATCAAAATTTGGTGAATTTGACATTTGAAGATTTCTGATTTCTTGT 240
QY 241 CACTTTAATTACTTGAGAAATTTAATGAATTTGGAGTCAGCAGAACTCAAAATAGGCAAA 300
|||||
Db 241 CACTTTAATTACTTGAGAAATTTAATGAATTTGGAGTCAGCAGAACTCAAAATAGGCAAA 300
QY 301 TACTGTGCTGGGTTGCAATGAACCAATTCATTAATGAATCAAAAGGCAATGAATACACA 360
|||||
Db 301 TACTGTGCTGGGTTGCAATGAACCAATTCATTAATGAATCAAAAGGCAATGAATACACA 360
QY 361 TTGCTGTTCATGAGTGAATCATGTTTCTGGACGGGATTTTGGCCCTCATCTCTGTT 420
|||||
Db 361 TTGCTGTTCATGAGTGAATCATGTTTCTGGACGGGATTTTGGCCCTCATCTCTGTT 420
QY 421 ATAGATAAACAAGATCTAATTAATTTGTTGGACACTGCAATCAAAATTTTGGAACTCTAG 480
|||||
Db 421 ATAGATAAACAAGATCTAATTAATTTGTTGGACACTGCAATCAAAATTTTGGAACTCTAG 480
QY 481 TTCACTTAAGTACTGCCAGCTGGTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
|||||
Db 481 TTCACTTAAGTACTGCCAGCTGGTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
QY 541 CCTCATGATATAGAGATTTCTCGCCATTTGTCATGGCTGGTGTGTCATGCGAGAGTAGTG 600
|||||
Db 541 CCTCATGATATAGAGATTTCTCGCCATTTGTCATGGCTGGTGTGTCATGCGAGAGTAGTG 600
QY 601 TCAACACAGCTTGGCGGCAATCACTGTTGTAATTAAGTAAAGTATTCCTCTATTAAGAA 660
|||||
Db 601 TCAACACAGCTTGGCGGCAATCACTGTTGTAATTAAGTAAAGTATTCCTCTATTAAGAA 660
QY 661 AGTTCTTTGGCTTAACAACCTGATCTGTGGGACACTTATCTACAAGTCTTTTACA 720
|||||
Db 661 AGTTCTTTGGCTTAACAACCTGATCTGTGGGACACTTATCTACAAGTCTTTTACA 720
QY 721 TTTAAGACAAGTGGATTTATGGAACACTGGGGATGGATCTGGTGTGATCGGGATCTCT 780
|||||
Db 721 TTTAAGACAAGTGGATTTATGGAACACTGGGGATGGATCTGGTGTGATCGGGATCTCT 780
QY 781 CAAATTAACAGCATCTGCTGTGGAGTGGACTGACACACAGGGCAAGAACAGTCTGG 840
|||||
Db 781 CAAATTAACAGCATCTGCTGTGGAGTGGACTGACACACAGGGCAAGAACAGTCTGG 840
QY 841 AAACCCAAAAAGCCAGGCTGAAAAACCTGGACCCCTTGGGCTGCTTTTGGCCACTGAT 900
|||||
Db 841 AAACCCAAAAAGCCAGGCTGAAAAACCTGGACCCCTTGGGCTGCTTTTGGCCACTGAT 900
QY 901 GAATACCAAGTGTACAAATAGATTTGAATAGGAAAGAAATAACAGGCATTTATAACC 960
|||||
Db 901 GAATACCAAGTGTACAAATAGATTTGAATAGGAAAGAAATAACAGGCATTTATAACC 960
QY 961 ACTGGATCCACCATGTTGGAGCACAATTTACTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
|||||
Db 961 ACTGGATCCACCATGTTGGAGCACAATTTACTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
QY 1021 GATGATGGCAGAAATGGACTGTGTACAGAGAGCTGTTGGTGGAGCAAGATAGATATTT 1080
|||||
Db 1021 GATGATGGCAGAAATGGACTGTGTACAGAGAGCTGTTGGTGGAGCAAGATAGATATTT 1080
QY 1081 CAAGGAAACAAGATTTATCACCAGGATGCGTAATAACTTTTGGCCACAATTTATGCA 1140
|||||
Db 1081 CAAGGAAACAAGATTTATCACCAGGATGCGCGTAATAACTTTTGGCCACAATTTATGCA 1140
QY 1141 CGTTTATTAGAGTGAATCTCTCCCAATGGCAGCAGAAAAATTCCTCATGAAATGGAGCTG 1200
|||||
Db 1141 CGTTTATTAGAGTGAATCTCTCCCAATGGCAGCAGAAAAATTCCTCATGAAATGGAGCTG 1200
QY 1201 CTCGGATGTCAGTTTATCTTAAGGTCTGCTCTCCAAAACTTACTCAACCTCCACCTCCT 1260
|||||
Db 1201 CTCGGATGTCAGTTTATCTTAAGGTCTGCTCTCCAAAACTTACTCAACCTCCACCTCCT 1260
QY 1261 CGGAACAGCANTGACCTCAAAAACACTACAGCCCTCCAAAAATAGCCAAAGCTCGTGC 1320

Db 1261 CGGAACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAAATAGCCAAAGTCTGTGC 1320
QY 1321 CCAAAATTTACCAACCACTACAACTCCAGTAGCAATGAATTTCTCTGCACAGACAGAA 1380
|||||
Db 1321 CCAAAATTTACCAACCACTACAACTCCAGTAGCAATGAATTTCTCTGCACAGACAGAA 1380
QY 1381 CAAACAACCTGCGCTCTGATATCAGAAATACTACCGTAACTCCAATGTAAACCAAGAT 1440
|||||
Db 1381 CAAACAACCTGCGCTCTGATATCAGAAATACTACCGTAACTCCAATGTAAACCAAGAT 1440
QY 1441 GTAGCGCTGGCTGCGAGTTCTTCTCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
|||||
Db 1441 GTAGCGCTGGCTGCGAGTTCTTCTCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1501 ATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
|||||
Db 1501 ATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
QY 1561 TTACCTTTACTGGGACCGGCGAGTTGGTGGAAAGGAATGAAGCAGTTTCTTCTGCAAAA 1620
|||||
Db 1561 TTACCTTTACTGGGACCGGCGAGTTGGTGGAAAGGAATGAAGCAGTTTCTTCTGCAAAA 1620
QY 1621 GCAGTGGACCATGAGGAAACCCAGTTGCTATAGCAGCAGGCAAGTTAATCACTGAGT 1680
|||||
Db 1621 GCAGTGGACCATGAGGAAACCCAGTTGCTATAGCAGCAGGCAAGTTAATCACTGAGT 1680
QY 1681 CCAAGAGAAGTCAACACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
|||||
Db 1681 CCAAGAGAAGTCAACACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
QY 1741 GGAGGAATTTGGTACACTTCAATCAAGATCTACCTTTTAAACCAGGAAGGAAGAA 1800
|||||
Db 1741 GGAGGAATTTGGTACACTTCAATCAAGATCTACCTTTTAAACCAGGAAGGAAGAA 1800
QY 1801 GCAGGCTATGACAGACCTAGATCTTACAACTCACCAGGCGAGGAAGTTTATCATGCTAT 1860
|||||
Db 1801 GCAGGCTATGACAGACCTAGATCTTACAACTCACCAGGCGAGGAAGTTTATCATGCTAT 1860
QY 1861 GCTGAACCACTCCCAATTTACGGGCTGAGTATGCAACCCCAATCATGACATGTCA 1920
|||||
Db 1861 GCTGAACCACTCCCAATTTACGGGCTGAGTATGCAACCCCAATCATGACATGTCA 1920
QY 1921 GGGCAGCCCACTCACTTCACTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1980
|||||
Db 1921 GGGCAGCCCACTCACTTCACTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1980
QY 1981 CAACTCCCACTCACTTCACTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2040
|||||
Db 1981 CAACTCCCACTCACTTCACTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2040
QY 2041 TCAGCCAGCCCACTCACTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2100
|||||
Db 2041 TCAGCCAGCCCACTCACTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2100
QY 2101 GAATTTGGTGTACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2160
|||||
Db 2101 GAATTTGGTGTACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2160
QY 2161 GAATGTGATGTTTTTAAAGAAATCTTTTGA 2190
|||||
Db 2161 GAATGTGATGTTTTTAAAGAAATCTTTTGA 2190

RESULT 2
US-09-879-536-480/c
; Sequence 480, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.

Qy 1714 TCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATTGTTGGTACACATTCATCAAGATCT 1773


```

; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2145
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate nucleotide sequence
; NAME/KEY: misc.feature
; LOCATION: (1)-(2145)
; OTHER INFORMATION: n = A,T,C or G
;
US-10-003-132-12

Query Match 13.4%; Score 293.2; DB 9; Length 2145;
Best Local Similarity 35.1%; Pred. No. 4.7e-78;
Matches 429; Conservative 229; Mismatches 522; Indels 42; Gaps:

Qy 10 TTCTCTCGCTCTTACTTTCCTGCTCTGCTGCTCGAGAGCGCTGGAGCCGACGACAGT 69
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 YTYTNGCNVTNTYNTNGCNGTWNWNGCCNYTNMGNYTNCARCGNGARGARYTNGN 114
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 70 GATGGATGGGACACACTGACTAGCCCTGAGAGTGGAGACCTTACATCCATAAACTAC 129
||: || || || || : || : || : || : || : || : || : || : || : || : || :
Db 115 GAYGNGTYGGNCAYTYTNGTACATYARGAYWNGGNACNATGACNWSNARAAYTAY 174
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 130 CCACAGACCTATCCCAACAGCACTGTTTGTGAATGGGAGATCGTGTAAAGATGGGAGAG 189
: || || || || || : || || || || : || : || : || : || : || : || : || :
Db 175 CCNGGNACNTAYCCNAYCAYACNGTNTGYGARAARACNATHACNTCCNARGGNAAR 234
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 190 AGATTTGCGCATCAAAATTTGGTGACTTTGACATTTGAAGATTTCTGATTTCTCTACTTTAAT 249
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 235 MGNVTNATHVTNMGNTWNGNGAYTYNGAYATHGAR--WSNCARACNTGYGCNWSNGAY 291
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 250 TACTTGGAGATTTATATGGAAATGGAGTCAGCAGACACTGAAATAGGCAAACTACTGTGT 309
||: || || : || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 TAYTNTYNTY-----ACNWSNWSNGAYCARTAYGNGCNCNTAYTGY 336
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 310 CTGGGTGTGCAAAATGAACATTCAAATGTAATCAAAAGGCAATGAAATCACATTGCTGTTC 369
: || : || : || : || : || : || : || : || : || : || : || : || : || :
Db 337 GGNWSNATGACNGTNCNARGARYNTYNTNAAACNWSNGARGTNACNGTNGMNTTY 396
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 370 ATGATGGAAATCCATGTTTCTGGACGGGATTTTGGCCTCATCTGTTATAGATAAA 429
: : : || : || : || : || : || : || : || : || : || : || : || : || :
Db 397 GARWSNGGNWSNCAYATHWSNGCNMGNGNTYTYTNTNACNTAYGCNWSNWSNGAYCAY 456
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 430 CAAGATCTAATTACTTTTGGACACTGCATCCAAATTTTTTGGAACTGAGTTCAGTAAG 489
| ||: || || || || || : || : || : || : || : || : || : || : || : || :
Db 457 CCNGAYTYNATHACNTGYTNGARMGNCNWSNCAYTAYTYNAAACNAGNARTAYWSNAAAR 516
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 490 TACTGCCACGCTGGTGTCTGCTTCTTTTGTGAGATATCTCGAACAAATTCCTCATGGA 549
| : ||: || || || || : || : || : || : || : || : || : || : || : || :
Db 517 TTYTGCCNGCNGNTGYMNGAYGTNCGNGNGAYATHWSNGGNAAATFGTNGAYGN 576
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 550 TATAGAGATTCCTCGCATTTGCATGGCTGGTGTGCTGACGAGCTAGTGTCAAAACAG 609
||: || || : : : || || || : || || || : || || || : || || || : || || :
Db 577 TAYMNGAYACNWSNNTYNTGYAARGCNGCNATHCAYCCNGNATHATGCGNGAYGAR 636
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 610 TTGGGGCGGCAATCAGTGTGTAATTAGTAAAGGATATCCCTATTATGAAAGTCTCTTG 669
: || || || || || : || : || || || : || : || || || : || : || || || :
Db 637 YTINGGNGNCARATHWSNGTNYTNCARMGNAARGGNATHWSNMGNTAYGARGGNATHYN 696
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 670 GCTAACACGTCACATCTGCTGGGACACTTATCTACAAGTCTTTTACATTTTAGACA 729
|| || : || : || : || : || : || : || : || : || : || : || : || :
Db 697 GCNAAAYGGNGTNTNWSNMGNGAYGGNWSNYTNWSNGAYAARMGNTTYTNTTYACNWSN 756
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 730 AGTGGATGTTATGAACACTGGGATGGAGTCTGGTCTGTGATCCGGATCTCCTCAATAACA 789
| : || || || : || : || : || : || : || : || : || : || : || : || :
Db 757 AAYGNGTYWSNMGNSNTYTNWSNTTYGARCCNGAYGG-----NCARATHMGN 804
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 790 GCATCATCTGTGTGAGTGGAGTCTGACCACACAGGGGCAAGAGAACACAGTTGGAAACCCAAA 849

```

Db 805 GCNNWSNWSNTGGCARWSNGTNAAYGARWSNGNGAYCARGTNCATTTGGWSNCCNGN 864
Qy 850 AAAGCCAGGCTGAAAAAACCTGGACCGCCTTTGGGCTGGCTTTGGCCACTGATCAATACC-- 907
Db 865 CARCNGNNTNCARGAYCARGGCCNWSNTGGCWNWSNGNGAYWWSNWSNAAAYACAY 924
Qy 908 -----AGTGGTTACAATAGATTGAATAGGAAAAAGAAAAATAACAGGCATTATA 957
Db 925 AARCCNMGARCTGGTNGARATHGAYTYTNGNGARAARAARAARATHACNGNATHMGN 984
Qy 958 ACCACTGCATCCACCTGGTGGAGCAACATTACTATGTGCTCGCTACAGAACTCCTGTAC 1017
Db 985 ACNACNGNWSNACNARWSNAAYTYAYTTAYGTNNAARWSNTTGTNATGNAATY 1044
Qy 1018 AGTGATGATGGCAGAAATGGACTGTGTACAGAGAGCCTGTGTGGAGCAAGATAAGATA 1077
Db 1045 AARAAYAAAYWSNAARTGTAARACNTAYAAARGGNATHGTNAAYAYGARGARAARGTN 1104
Qy 1078 TTCAAGGAACAAGATTTATCACAGGATGTGCGTAATAACTTTTTTCCCAACCAATATT 1137
Db 1105 TTYCARGGNAAYWSNAAYTYTNGAYCNGTNCARAAYAYTYTATHTCCNCCNATHGTN 1164
Qy 1138 GCAGTTTTATGAGTGAATCTCCCAATGGCAGCAGAAAAATTTGCCATGAAATGGAG 1197
Db 1165 GCNMGNTAYGTNMGNTGNTGCCNARACNTGGCAYCARGMNGNTHGCTNNAARGTNGAR 1224
Qy 1198 CTGCTCGGATGTCAGTTTATTC 1219
Db 1225 YTNATHGNTGYCARATHACNC 1246

RESULT 6
US-09-864-761-12562
: Sequence 12562, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharron G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
: FILE REFERENCE: Aemlica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661

RESULT 6
US-09-864-761-12562
; Sequence 12562, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GENE EXPRESSION ANALYSIS BY MICROARRAY


```
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (76)...(2223)
; US-10-003-132-1

Query Match      10.2%; Score 224.2; DB 9; Length 3151;
Best Local Similarity 52.0%; Pred. No. 4.6e-57;
Matches 607; Conservative 0; Mismatches 518; Indels 42; Gaps 3;

QY 62 AGCAAGGTGATGGATGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTTACATCCA 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 AGCTGGGTGATGGCTGGACACCTAGTACTATCAGAGATAGTGGCACAATGACATCTA 241

QY 122 TAACTACCCACAGACCTATCCCAACAGCACTGTTTGTGAATGGAGATCCGCTGTAAGA 181
   || || || || || || || || || || || || || || || || || || || || ||
Db 242 AGAATTATCCCGGACCTACCCCAATCACACTGTTTGGGAAAGACAATTTACAGTACC 301

QY 182 TGGGAGAGAGTTCGCATCAATTTGGTGACTTGGACATTTGAAGATTTCTGATTTCTGTC 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 302 AGGGGAAAGACTGATTTCTGAGTTGGGAGATTTGGATTCGAA ----- 345

QY 242 ACTTTAATTACTTGAAGATTTATATGAATTTGGAGTGCAGAGAACTGAAATAGGCAAA 301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 346 --TCCAGACCTGTCTCTGACTATCTCTTCCACAGCTCTTCAGATCAATATGTC 403

QY 302 ACTGTGGTCTGGGTTGCAATGAACCAATCAATTTGAATCAAAAGGCAATGAAATACAT 361
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 404 CATAGTGTGAAGTATGACTGTCTCCAAAGAACTCTTGTGAACAACAAGTGAAGTAA 463

QY 362 TCCTGTTCATGAGTGAATCCATGTTCTGGACGGGATTTTGGCCTCATACTCTGTTA 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 464 TCCGCTTTGAGAGTGGATCCCAATCTCTGCGCGGGTCTTGTGACCTATGCGAGCA 523

QY 422 TAGATAACAAGATCAATTAATCTTTGGACACTGCATCCCAATTTTGGAACTCTGAGT 481
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 524 GCGACCATCCAGATTTATACATGTTTGGACGAGCTAGCCATTATTGAGACAGAT 583

QY 482 TCAGTAAGTACTGCGCCAGCTGGTCTGCTGCTTCTTTGCTGAGATATCTGGAACAATTC 541
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 584 ACAGCAAAATCTGCCAGCTGGTTGTAGAGACGTAGCAGGAGACATTTCTGGGAATATGG 643

QY 542 CTCATGGATATAGAGATTCCTCGCCATTTGTCATGGCTGGTGTGATCGCAGGAGTAGTGT 601
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 644 TAGATGGATATAGAGATACCTCTTTATTGTGCAAGAGTGGCCATCCATCGAGGAATAAT 703

QY 602 CAACACAGCTTGGCGGCCAAATCACTGTTGTAATAGTAAGGTATTCCTATTATGAAA 661
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 704 CTGATGAAGTACTGAGTGGCCAGATCACTGCTTCAGCGCAAGAGGATCACTCATATGAG 763

QY 662 GTTCTTTGGCTAAACAGCTACATCTGTGTGGGACACTTAATCTACAAGTCTTTTACAT 721
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 764 GGATTTCTGGCCAAATGGTGTCTTTCCAGGGATGGTTCCTGTGCAGCAAGCGATTCTCT 823

QY 722 TTAACACAAGTGGATTTATGGAACACTGGGATGGATCTGTTGTGATCGCGGATCTCTC 781
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 824 TTAACCTCAANTGGTTGCGACAGATCTTGGATTTTGAACCTGACGGG-----C 871

QY 782 AATAACAGCATCATCTGCTGTGGAGTGACTGACACAGCGGCAAGAGACAGTTGA 841
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 872 AATCAGAGCTTCTTCTCATCGGAGTGGTCAATGAGAGTGGAGACCAAGTTCATGTT 931

QY 842 AACCCCAAAAGCCAGGCTGAAAAACCTGGACCGCTTGGCTGCTTTTGGCACTGATG 901
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
```

217	ATGGCTGTGGGCACATAGTGAACCTCTCAGGACAGTGGCCAAATGACATCTAAGAANTATC	276
131	CACAGACCTATCCCAACAGCACTGTTTGTGAATGGGAGATCCGTGTAAAGATGGGAGAGA	190
277	CAGGGACTTACCCCAATTACACTGTGTGTGAAAGAATCATCAGATCCCAAGGGGAAGA	336
191	GAGTTCGCATCAAAATTTGGTGACTTTGCAGTATGAAGATTCTGATTTCTGTCACTTTAAT	250
337	GACTTATCTGAGGTTGGGAGATTTGAACATTTGA-----GTCCAAGACCTGC	383
251	ACTTTGAGAATTTAATAAGGAATGGAGTTCAGCAGCAACTGAAATAGGCAAAATACTGTGGTC	310
384	GCCTTCGACTATCTCCTTCACGAGTGCACAGATCAGTATGGTCCATATTGTGGGACT	443
311	TGGGGTTGCAATGAACCATTCAAATTGAATCAAAAGCAATGAATCAACATGCTGCTTCA	370
444	TGGGCTGTTCCTCCAAAGAACTCCGGCTGAATCA-----AACGSAAGTACTGTCTCTTCA	498
371	TGAGTGGAACTCAATGTTCTTGGAGCGGATTTTTTGGCTCATACTCTGTTATAGATAAAC	430
499	AGAGTGGATCTCACATTTCTGGCCGGGGCTTCTGTGACCTACGCCAGCAGTGACCATC	558
431	AAGATCTAAATPACTTTGTGGACACTGCATCCAAATTTTTTGGAAACCTGAGTTCAGTAAGT	490
559	CAGATTTAATAAACCTGTTTGGAAAGGCGACCATTAATTTTCGAGGAAAAATACAGCAAT	618
491	ACTGCCACGCTGGTGTCTGCTCTTTTCTCAGAGATATCTGAAACAATTCCTCATGGAT	550
619	TCGCCACGCTGGCTGTGAGACATACGACGAGATATTTCTGGGAATCAAAAGATGGTT	678
551	ATAGAGATTCCTCGCCCATTTGTGATGGCTGGTGTGCATGCAGGAGTAGTGTCAAAACAGT	610
679	ACAGAGATACCTCTTTATTGTGCAAGCTGCCATCCACGCGGGATCATCACAGATGAAC	738
611	TGGGGGGCCAAATCAGTGTGTGTAAATAGTAAGGATATTCCTATTATGAAGTTCCTTGG	670
739	TAGTGGCCACATCAACTGCTTTCAGAGCAAAAGGATTAAGTCTACTATGAAGSACTCCTGG	798
671	CTAACAAAGTCACATCTGTGGTGGGACACTTATCTACAAGTCTTTTATACATTTAAGACAA	730
799	CCAATGGCGTCTCTCCGGCATGGTCTTTGTTCGGAAGAGCGATTTCTTTTACAACC	858
731	GTGGATGTTATPAGAACACTGGGGATGGAGTCTGGTGTGATC	771
859	CAGCAATGAATATTACAACTGTGGCGATTCCATCAGTGATC	899

```

RESULT 11
US-09-823-038A-43
: Sequence 43, Application US/09823038A
: Patent No. US20020058335A1
: GENERAL INFORMATION:
: APPLICANT: Strachan, Lorna
: APPLICANT: Sleeman, Matthew
: APPLICANT: Abernethy, Nevin
: APPLICANT: Onrust, Rene
: APPLICANT: Kumble, Anand
: APPLICANT: Murison, Greg
: TITLE OF INVENTION: Compositions Isolated From Stromal Cells
: TITLE OF INVENTION: and Methods For Their Use
: FILE REFERENCE: 11000.1037C3
: CURRENT APPLICATION NUMBER: US/09/823, 038A
: CURRENT FILING DATE: 2001-07-09
: NUMBER OF SEQ ID NOS: 61
: SOFTWARE: FastSeq for Windows Version 4.0

```

Query Match 6.8%; Score 148.2; DB 10; Length 1871;
Best Local Similarity 51.9%; Pred. No. 3.2e-34;

QY	613	AAACCCAAAAGCCAGCGCTGAAAAACCTCGAC-----CGCCTTGGGCTGCTTTTGGC	666
Db	6379	GAACCTCTCCGTGGCCGCTCTGAATGCCAGGACGCTGTGAATCGCTGGCAAGCCAAGGCA	6438
QY	667	ACTGATGAATACCACTGGTTACAAATAGATTAAGGAAAAGAAATACAGCGCAAT	726
Db	6439	AACAACAATAAGCAGTGGCTAGAAATTGATCTACTCAAGATCAAGAAAGATTAAGCGGAATT	6498
QY	727	ATAACCACTGGATCCACCATGGTGGAGCACAAATTACTATGTGTGCTGCTACAGAAATCTTG	786
Db	6499	ATACACAGGGCTGCAAGTCTCTGCTCTGAAATGTATGTTAAAGACTATACCATCCAC	6558
QY	787	TACAGTGATGATGGCGAATAATGGACTGTGTACAGAGAGCCTGGTGGGAGCAAGATAAG	846
Db	6559	TACAGTGACAGGGAGTGGAAATGGAACCATACAGGCTGAAATCCTCATGGTGGACAAG	6618
QY	847	ATATTTCAAGGAAACAAGAATTATCACCAAGATGCGGTAAATACTTTTGGCCACCAAT	906
Db	6619	ATTTTGAAGGAAATACTAATAACCAAGAGCATGTGAAGAACTTTTTCACACCCCCCAATC	6678
QY	907	ATTGCACGTTTATTAGAGTGAATCCTACCAATGGCAGCAGAAAAATGCCATGAAAAATG	966
Db	6679	ATTTCCAGGTTTATCCGTGTCATTCCTTAAACATGGAATCAAAGTATTACACTTCGCGCTG	6738
QY	967	GAGCTGCTCGGATGTCAGTTT	987
Db	6739	GAACTCTTTGGGCTGTGATATT	6759

RESULT 15
US-10-115-563-26
; Sequence 26, Application US/10115563
; Publication No. US20030008307A1
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H
; Greengard, Judith S
; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTAT
; AND COMPOSITIONS THEREOF

NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
Patent Counsel
STREET: 10666 No. US20030008307Alth Torrey Pines Road, TPC 8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/115,563
FILING DATE: 02-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/410,488
FILING DATE: 24-MAR-1995

ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: 44,9.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 6925 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

	Query Match	4.48;	Score 86.6;	DB 9;	Length 6925;
	Best Local Similarity	51.7%;	Pred. No. 2.4e-15;		
	Matches 259;	Conservative 0;	Mismatches 224;	Indels 18;	Gaps 2;
Qy	493	TTTAAGACAAGTGGATGTTATTGAACACTGGGGATGGAGTCTGTGTGTGATCGCGGATCCT	552		
Db	6271	TGTGAGGTAAATGGATGTTCCACACCCTGGTGATGAAAATGCAAGATAGAAAAAAG	6330		
Qy	553	CAAATAACAGCATCTGTGTGGAGTGGACTGACACACAGCGGCCAAGAACAAGTTGG	612		
Db	6331	CAAATCACAGCTTCCTC-----GTTTAAGAAATCTGGTGGGAGATTACTGG	6378		

Wed Jan 22 14:57:21 2003

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-10-115-563-26

Query Match      4.4%; Score 86.6; DB 9; Length 6925;
Best Local Similarity 51.7%; Pred. No. 2.4e-15;
Matches 259; Conservative 0; Mismatches 224; Indels 18; Gaps 2;

Qy 493 TTAAAGCAAGTGGATGTTATGAACACATGGGATGGAGTCTGTGTGATCGGGATCCT 552
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6271 TGTGAGTAAATGGATGTTCCACACCCCTGGGTATGGAATGGAAGATAGAAAACAAG 6330
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 553 CAAATAACACATCATCTGTGCTGGAGTGGACTGACACACAGGCGCAAGAGAACAGTTGG 612
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6331 CAAATCACACTTCTTC-----GTTTAAAGAAATCTTTGGTGGGAGATTACTGG 6378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 613 AAACCCAAAAAGCCAGGCTGAAAAAACCTGGAC-----CGCCTTGGGCTGCTTTTGCC 666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6379 GAACCTTCCGTGCCCTCTGAATGCCAGGACGCTGTGAATGCTGGCAAGCCCAAGGCA 6438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 667 ACTGATGAATACCACTGGTTACAAATAGATTGAATAAGGAAAGAAATAACAGGCATT 726
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6439 AACAACAATAAGCAGTGGCTAGAAATTTGATCTACTCAAGATCAAGAAGATAACGGCAAT 6498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 727 ATACCACCTGGATCCACCATGGTGGAGCACAAATTACTATGTCTGTGCTTACAGAAATCCTG 786
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6499 ATAACACAGGCTGCAAGTCTCTCTCTCTGAATGTATGTAAAGAGCTATACCATCCAC 6558
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 787 TACAGTGAATGGGCAAGAAATGACTGTGTACAGAGAGCCGTGTGTGGAGCAAGATAAG 846
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6559 TACAGTGAGCAGGAGTGGAAATGGAATGGAACCAACATACAGGCTGAAATCCTCCATGGTGGACAAG 6618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 847 ATATTTCAAGGAACAAGATTATCACCAGGATGTCGTAATAACTTTTCCACCACCAATT 906
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6619 ATTTTGAAGGAATATCTAATACCAGGACATGTGAAGAACTTTTCAACCCCCCAATC 6678
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 907 ATTGCACGTTTTATTAGAGTGAATCCCTACCAATGGCAGCAGAAAAATGGCATGAAAATG 966
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6679 ATTTCCAGGTTTATCCGTTGTCATTCCTTAAACATGGAATCAAAAGTATTACACTTCGCCTG 6738
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 967 GAGCTGCTCGGATGTCAGTTT 987
    ||||| ||||| ||||| |||||
Db 6739 GAACCTTTTGGCTGTGATATT 6759
    ||||| ||||| ||||| |||||
```

Search completed: January 21, 2003, 09:30:57
Job time : 125.14 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:51:36 ; Search time 8.43994 Seconds
(without alignments)
1717.007 Million cell updates/sec

Title: US-10-060-830-3
Perfect score: 3888
Sequence: 1 MPFLLLLVLLLLLEDAGA.....TQVSGAGRDGCDVKEIL 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1042	26.8	715	9 US-10-003-132-2	Sequence 2, Appli
2	575.5	14.8	503	10 US-10-003-132-4	Sequence 4, Appli
3	574.5	14.8	503	10 US-09-823-038A-51	Sequence 51, Appli
4	465	12.0	458	9 US-10-003-132-6	Sequence 6, Appli
5	412	10.6	75	10 US-09-864-761-45371	Sequence 45371, A
6	400.5	10.3	923	12 US-10-104-440-2	Sequence 2, Appli
7	395.5	10.2	931	12 US-10-104-440-4	Sequence 4, Appli
8	347	8.9	2224	9 US-10-115-563-14	Sequence 14, Appli
9	327.5	8.4	1431	12 US-10-095-718-4	Sequence 4, Appli
10	308.5	7.9	1438	12 US-10-006-091-1	Sequence 1, Appli
11	308.5	7.9	1438	12 US-10-047-257-1	Sequence 1, Appli
12	308.5	7.9	1471	12 US-10-095-718-2	Sequence 2, Appli
13	308.5	7.9	2332	9 US-09-957-641-2	Sequence 2, Appli
14	220.5	5.7	756	9 US-09-992-598-62	Sequence 62, Appli
15	220.5	5.7	756	9 US-09-989-293A-62	Sequence 62, Appli
16	220.5	5.7	756	9 US-09-989-735-62	Sequence 62, Appli
17	220.5	5.7	756	9 US-09-990-444-62	Sequence 62, Appli
18	220.5	5.7	756	9 US-09-989-730-62	Sequence 62, Appli
19	220.5	5.7	756	9 US-09-990-436-62	Sequence 62, Appli

20	220.5	5.7	756	9 US-09-991-181-62	Sequence 62, Appli
21	220.5	5.7	756	9 US-09-993-687-62	Sequence 62, Appli
22	220.5	5.7	756	9 US-09-989-734-62	Sequence 62, Appli
23	220.5	5.7	756	9 US-10-028-072-392	Sequence 392, Appli
24	220.5	5.7	756	9 US-09-997-653-62	Sequence 62, Appli
25	220.5	5.7	756	10 US-09-989-723-62	Sequence 62, Appli
26	220.5	5.7	756	10 US-09-989-723-62	Sequence 62, Appli
27	220.5	5.7	756	10 US-09-989-279-62	Sequence 62, Appli
28	220.5	5.7	756	10 US-09-989-721-62	Sequence 62, Appli
29	220.5	5.7	756	10 US-09-989-731-62	Sequence 62, Appli
30	220.5	5.7	756	10 US-09-989-732-62	Sequence 62, Appli
31	220.5	5.7	756	10 US-09-991-073-62	Sequence 62, Appli
32	220.5	5.7	756	10 US-09-990-442-62	Sequence 62, Appli
33	220.5	5.7	756	10 US-09-991-163-62	Sequence 62, Appli
34	220.5	5.7	756	10 US-09-993-604-62	Sequence 62, Appli
35	220.5	5.7	756	10 US-09-990-456-62	Sequence 62, Appli
36	220.5	5.7	756	10 US-09-989-721-62	Sequence 62, Appli
37	220.5	5.7	764	10 US-09-900-518A-2	Sequence 2, Appli
38	200.5	5.2	855	10 US-09-771-161A-196	Sequence 196, Appli
39	198	5.1	277	10 US-09-799-118-2	Sequence 2, Appli
40	198	5.1	734	9 US-10-174-590-458	Sequence 458, Appli
41	198	5.1	734	9 US-10-176-758-458	Sequence 458, Appli
42	198	5.1	734	12 US-10-052-586-458	Sequence 458, Appli
43	194	5.0	144	9 US-09-886-429-4	Sequence 4, Appli
44	194	5.0	144	9 US-10-041-406-5	Sequence 5, Appli
45	191.5	4.9	986	10 US-09-285-385C-19	Sequence 19, Appli

ALIGNMENTS

RESULT 1
US-10-003-132-2
; Sequence 2, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: SHOEEMAKER, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-132-2

Query Match : 26.8% ; Score 1042; DB 9; Length 715;

Best Local Similarity 36.5%; Pred. No. 5.5e-76;

Matches 264; Conservative 113; Mismatches 238; Indels 108; Gaps 24;

QY	5	LLLLLVLLLLLEDAGAAGCGGCHTVLGPESGTLTSINYPQYPNSTVCWEIRVKMGER	64
DB	20	LALLAVSAPLRQLQAEELGDCGHLVYDSTGWTSKNYPGTYPNHTVCEKTIITVPKGR	79
QY	65	VRKFGDFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGIGLQMNHISKSGNEITLLPM	124
DB	80	LILRLGDLIE-SQTCASDYLFTS-----SSDQGYPCG-SMTVPKRELLNTSEVTFRF	133
QY	125	SGHVSGRFLASYSVIDKQDLITCLDTASNFLEPFESKYCPAGCLLPFAEISGTIPHCY	184
DB	134	SGSHISGRGFLTYASSDHPDLITCLERASHYLKTEYSKFCPAGCRDVGDSGNMVDGY	193
QY	185	RUSSPLCMAGVHAGVSVNTLGGQISVVISKGPYYESSLANNVTSVVGHLSLSTSLFFKTS	244
DB	194	RTSLACKAAIHAGIITADELGGQISVLQKRGISRYEGILANGVLSDGSLDKRFLFTSN	253


```

Db      8 P S V L A L F A V C A B L R Q A E L G D G C G H I V T S Q D S G T W T S K N T P G T V P N T V T V C E K I I T V P K 67
QY      62 G E R V R I K F G D F I E D S D S F N F N T L R I Y N G I G V S K R T E I G K Y C G L G Q M N H S I B S K G N E I T L 121
Db      68 G K R L I R L G D L N I E - S K T C A S D Y L L F S S - - - A T D Q Y G P Y C G - S W A V P K E L R L N S N E V T V 121
QY      122 L F M S G I H V S G R F L A S Y S I D K O D L I T C L D T A S N F L E P E S F K Y C P A G C L L P F A E T S G T I P 181
Db      122 L F K S G S H I S G R F L L Y A S D H P D L I T C L E R G S H Y P E E K Y S K P C P A G C R D I A R D I S G N T K 181
QY      182 H G Y R D S P L C M A G V H A G V Y S N T L G G Q I S V V I S K I G P Y Y E S S L A N N Y T S V V G H L S T S L F T 241
Db      182 D G Y R D S L L C A A I H A G I T D E L G G H I N L L Q S K G I S H Y E G L L A N G V L S R H G S L S E K R F L F 241
QY      242 K T S G C Y G T L G M E S G V I A D P Q I T A S S V L E W T D H T G Q E N S W K P K K A R L K K P C P W A A P A T D E 301
Db      242 T T - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 243
QY      302 Y O M L O L D L N K E K K I T G I T T G S T M V E H N Y V S A Y R I L Y S D G Q K T V Y R E P G V E Q D K I P Q 361
Db      244 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 246
QY      362 G N K D Y H O D V E R N N F L P P I A R I F V N P T O W O O K I A M K N E L L G C O F I P K R P P K L T Q P P P R 421
Db      247 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 246
QY      422 N S N D L K N T A P P K I A K G R A P K F T Q P L Q P R S S N E F P A Q T E Q T T A S P D I R N T V T P N V T K O V 481
Db      247 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 251
QY      482 A L A A V L P V L V M V I T L I L V C A W H R N R K K K T E G T Y D L P Y - - - - - W D R A G M W K M G K Q P L 537
Db      252 A I P S V I - F T A L L T G M G I A I C - - - - - R K R K K G N - - P Y V S A D A Q K T G C W K Q I K Y - - 298
QY      538 P A K A V D H E E T P - V R Y S S E - - - V N H L S P R E V T - - - T V L Q A D S A B Y A Q P L V G G I V G T U H Q R S 591
Db      299 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 348
QY      592 T F K P - - - - - B E - - - G K E A C Y A D L P V N S P Q E V Y H A Y A E P L I T Q E V A T P I I - - - - - 636
Db      349 T E R P M D P T E E V R N T E A S G H - - - - - Y D C P H R P G R H E Y A L P L T H S E P E V A T P I V E R H L L R 403
QY      637 - - - - - M D M S G H P T T S V G O P S T F K A T G N Q P P L V G V Y N T L L S R T D S C S A Q A Q Y D T P P K A G 692
Db      404 A H T F S T O S G Y R V P C - P R P T H E H S S G G F F P A T G A T Q V E S Y Q R P A S P K P V G G Y D K P - A A 461
QY      693 K P G L P A P D E L V Y Q V P O S T Q E V S G A G R D G 720
Db      462 S S F L D S R D - - - - - P A S O S M T S G G D P G 483

```

RESULTS. T. A.

US-10-003-132-6 ; Sequence 6, Application US/10003132
; Publication No. US20020192750A1

; GENERAL INFORMATION:

APPLICANT: FOX, Brian A.

; APPLICANT: Gao, Zeren

APPLICANT: Shoemaker, Kimberly E.

: TITLE OF INVENTION: NEUROPILIN HOMOLOG ZCUBS

: FILE REFERENCE: 00-62

; FILE REFERENCE: 00-62
: CURRENT APPLICATION NUMBER: US/10/003,132; CURRENT APPLICATION NUMBER: US/11-15-15
: CURRENT FILING DATE: 2001-11-15

; CURRENT FILING DATE: 2001-11-15
 : PRIOR APPLICATION NUMBER: US 60/249 004

; PRIOR APPLICATION NUMBER: US 60/
: PRIOR FILING DATE: 2000-11-15

; PRIOR FILING DATE: 2000-1-
 ; NUMBER OF SEQ ID NOS: 10

; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Ecdosec for Windows 3.0; SOFTWARE: Fast
: CEO ID NO C

; SEQ ID NO 6

; LENGTH: 45

```

; TYPE: PRT

```

; ORGANISM: MU

Query Match 12.0%; Score 465; DB 9; Length 458;

Best local similarity 22.5%: Pred. No. 1.1e-29;

Best local similarity	22.56, freq. no. 1.16	23,
Matches	168: Conservative	75: Mismatches 159: Indels 346: Gaps 21:

Qy	2	PLFLLILLVLLLLEDAGAOOGDCGHTVLGPESGTLTSTINYQTYPNSTVCWEIRVKM	61
Dd	8	PSVALLFAVCAPLRLQAEEELGGCGHIVTSQDSGTWTKSNYPCTTPNTVCEKIITVPK	67
Qy	62	GERVRIFGDFIEDSDSCHFNFLRIYNGTGVSRTEIGKYCGLGLOMNHSIESKGNEITL	121
Dd	68	GKRILRLGLDLNIE-SKTCAASY-----L	90
Qy	122	LFMGSHVSGRGFLASVYIDKODLTCLDTASFNLPEPFEFSKYCPACGLLPRAEISGTP	181
Dd	91	LFS-----SATDQDYDLITCLERGHSHYEEKYSKFCPACGRDIADGISGNTK	136
Qy	182	HGYRDSPLCMAGVHAGVSVNTLGGQISVVISIKGIPIYESSLANNTVSVVGHLSLTFE	241
Dd	137	DGYRDTSLUCAAIHAGIITDELGGHINLQSKGISHYEGLLANGVLSRHGSUSEKRFLE	196
Qy	242	KTSCYCYGLMGESGVIAADPOQTASSVLEWDHTGOENSMKKPKARLUKPPMPAAFPATDE	301
Dd	197	T-----PGM-----	198
Qy	302	YQWLQIDLNKKEKITGIITGSTWVEHYVYSAYRILYSDGQKVTVREPGVEQDK IFQ	361
Dd	199	-----PGM-----	201
Qy	362	GKNDYHQDVNRNFLPPIIARFIRVNPTWOQKIAMKMELGCOFIPKRPPKLTOPPPP	421
Dd	202	-----	201
Qy	422	NSNDLKNTTAPPKIAKGRAPKFTQPLQPRSNSEPPAQTEQTASPDIRNTVTTPNWTKD	481
Dd	202	-----NIT-TV	206
Qy	482	ALAAVLPVLMVLTLLILLVCAWHWRNRKKTKETGYDLPY---WDRAGMWKMKQFL	537
Dd	207	AIPSVI--FTALLTGMIFFAIC-----RRKKKGHN---PYVASADAQKTGCWKQIKY--	253
Qy	538	PAKAVDHEETP-VRYSSSE--VNHLSPREVT---TVLQADSABYAOPLVGGIVGTLQHRS	591
Dd	254	-----PFARHQSTTEFTISYDNEKEMTQKLDLITSDMADYQQPLMTGCTGVARKGS	303
Qy	592	TFKP-----GREAGVADLPYNPCQEVYHAYAEPLITGPEATPII-----	636
Dd	304	TRPMDDTEEVRYNVTCASGH-----YDCPHRGHRHEYALPLTHSEPETAIVERHLLR	358
Qy	637	----MDMSGHPITTSVGOPSTFRKATGNQPPLVGYNTLLSTRDSCSSAAQAYDTPKAG	692
Dd	359	APRESTOSGYBVDG-PRPTHKHSHSSGGFFEPATCATOVESORPASPKPVGGYDKP-AA	416

RESULT 5

US-09-864-761-45371

03-03-004701 45371 : sequence 45371. Application IIS/09864761

sequence 43371, application
Patent No. PS20020048763A1

Patent No. US2002004

APPLICANT: Penn. Sharron G.

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.

APPLICANT: HANZEL, DAVID R.
APPLICANT: HANZEL, DAVID R.

APPLICANT: HANZEL, David K.
APPLICANT: Chen, Wensheng

APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENE

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS OF A CELL LINE

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,7

; CURRENT FILING DATE: 2001-05

; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 45371
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
; OTHER INFORMATION: EST.HUMAN HIT: A1565996.1, EVALUATE 2.00e-40
; OTHER INFORMATION: SWISSPROT HIT: P98063, EVALUATE 1.00e-09
US-09-864-761-45371

Query Match 10.6%; Score 412; DB 10; Length 75;
Best Local Similarity 100.0%; Pred. No. 1.4e-26;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 24 DCCGHTVLGPSCGTLTSLNYQTPNVTVCWEIRVXMGKRVKFGDFIEDSDSCHFN 83
DB 1 DCCGHTVLGPSCGTLTSLNYQTPNVTVCWEIRVXMGKRVKFGDFIEDSDSCHFN 60
QY 84 YLRIYNGIGVSRTEI 98
DB 61 YLRIYNGIGVSRTEI 75

RESULT 6
US-10-104-440-2
; Sequence 2, Application US/10104440
; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Quan

; TITLE OF INVENTION: ANTAGONISTS OF NEUROPILIN RECEPTOR FUNCTION AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 48802 C
; CURRENT APPLICATION NUMBER: US/10/104,440
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/580,803
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/069,155
; PRIOR FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 60/069,687
; PRIOR FILING DATE: 1997-12-29
; PRIOR APPLICATION NUMBER: 60/078,541
; PRIOR FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 923
; TYPE: PRT
; ORGANISM: human
US-10-104-440-2

Query Match 10.3%; Score 400.5; DB 12; Length 923;
Best Local Similarity 28.1%; Pred. No. 4.8e-24;
Matches 130; Conservative 66; Mismatches 178; Indels 89; Gaps 18;

QY 6 LLLLVLLLLDAGAOOGDGGHTVLGPSCGTLTSLNYQTPNVTVCWEIRV-KMGER 64
DB 7 LILCAVLALVAPAGAFNRDKGDTIKIESPGYLSPGYPHSHFSEKCEWLIQADPPQR 66
QY 65 VRIKFG-DFDIEDSDSCHFNRYLRIYNGIGVSRTEIGVYCGGLQMNHHSIESKNEITLLF 123
DB 67 IMINFNPHFDLEDRD-CKYDYVEVDFGENENGHFRGKFC--GKIAPPVVSPPFLFIKF 123
QY 124 MSGIHVSGRGFLASYSVIDK-----QDLIT----- 148
DB 124 VSDYETHGAGFSIRYEIFKRGPECSQNYTPSPGVKSPGPEKYPNSLECTYIVFAPKMS 183
QY 149 --CLDTASNELEPESKYCPAGCLLPFAEISGTPHGYRDSPLCMAGVHAGVVS---N 202
DB 184 EILFEFSDLEPDSNP--PGMFCRYDRLE--IWDGPDVGP--HIGRYGQKTPGRIR 237
QY 203 TLGQISVWISKGIPYVESLAN-----NVTSVVGHLSLFTFKTSCTGTLGMSGYI 257
DB 238 SSSGILSMVF-----YDTSATAKEGSANYSVLOSSVED---PK---CMEALGMSGEI 286
QY 258 ADPOITASSVLEWTDHYGOENSWPKKARLAKKPGPPAAFAFATDEYQWLOIDLNKEKKTG 317
DB 287 HSDQITASS-----QYSTNWSAERSRLNYPENGWTPGEDSYREWIQVDLGLRFTYA 338
QY 318 IITG--STWVEHNYYVSAYRILYSDGQKWTVYREPGEQDKIFQGNKDYHQDVNNFL 375
DB 339 VGTGALSKETKKYIYKTYKIDVSSNGEDWITKEG--NKPVLFGQNTDVTVVAVFP 396
QY 376 PPIATAREIVNPTQWQOKIAMKMLLGCQIFKPGRPPLKTOPP 418
DB 397 KPLITREVRIPKATWETGISMRFEVYGC-----KITDYP 430

RESULT 7
US-10-104-440-4
; Sequence 4, Application US/10104440
; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Quan
; TITLE OF INVENTION: ANTAGONISTS OF NEUROPILIN RECEPTOR FUNCTION AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 48802 C
; CURRENT APPLICATION NUMBER: US/10/104,440
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/580,803
; PRIOR FILING DATE: 2000-05-30


```

; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

Query Match      8.4%  Score 327.5; DB 12; Length 1431;
Best Local Similarity 27.3%; Pred. No. 7.2e-18;
Matches 112; Conservative 58; Mismatches 144; Indels 97; Gaps 15;

QY 75 EDSDSCHFN-----YLRIN-----GIGVSRTE--IGKYCGLGL 106
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1050 ENIHSIHFGHVTVRKKEEYKMAVINYLPGVFETVEMLPFSQVGIWRIECLIGEHLOAGM 1109
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 107 QMNHIESKGNIEITLLEMSG-----IHVSCR-----GFLASYSVID---- 142
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1110 STLFVYSKCKQTLPGMASGHIRDFQITASQYQWAPKRLARLHYSGINAWTKDPFSW 1169
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 143 -KODLITCL--DNASNLEPEFSKYCPAGCLLPFAISGTPHGYRDSPLCMAGVHAG 198
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1170 IKVDLLAPMIHIGINTQAGQKFSVLAVSQFIIMYS-LDGNKWHYRGNS----- 1218
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 199 VVSNLGGQISVWISKGI-----PYESSLANVT--SVVGHLSLSTFTKTSYGYGT 249
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1219 --TGTLMVFFGNVDSSGKHNFPPIIAQYIRLHPHYSIRSLRMELLGCDNFSCSMP 1276
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 250 LGMESGVADPQITASSVLEWTDHTGOENSKWKPKKARLKP--PPWAAATDEYQWLQI 307
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1277 LGMESKAISDAQITASSYL-----SSMLATWSPQARLHOGRTNAPWAPNNPKEWLQV 1331
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 308 DLNKKKITGIITGTSMVHNHYVSAYRILYSDDGOKWTYVIRPGVEQDKIFOGNKDYH 367
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1332 DFRKTMKVGITTOGVKSLLSIMYVKEFLISSQDGHNWLFLQNG--KVKVFGNDRSS 1389
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 368 QDVRRNPLPIIARFIRVNPQWQOKTAMKMLLGCQFIPKGRPKLTQPP 418
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1390 TPVRNRLEPPLVARYRLHPQSWAHIALRLVLGCD-----TOOP 1430
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

RESULT 10
US-10-006-091-1
; Sequence 1, Application US/10006091
; Patent No. US20020102730A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.1
; CURRENT APPLICATION NUMBER: US/10/006,091
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-006-091-1

Query Match      7.9%  Score 308.5; DB 12; Length 1438;
Best Local Similarity 37.1%; Pred. No. 2.5e-16;
Matches 66; Conservative 31; Mismatches 72; Indels 9; Gaps 3;

```

```

QY 229 SVVGHLSLSTFTKTSYGYTLGMSGVIADPQITASSVLEWTDHTGOENSKWKPKARLK 288
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1263 SIRSTLRMELMGCDLNSCSMPLGMSKASDAQITASSYF-----TNMFATWSPSKARLH 1317
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 289 KPG--PPWAAATDEYQWLQIDLNKKKITGIITGTSMVHNHYVSAYRILYSDDGOKW 346
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1318 LQGRSNAWRPQVNNPKEWLQVDFQKTMKVGTGTQGVKSLLSIMYVKEFLISSQDGHQW 1377
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 347 TVYRPGVEQDKIFOGNKDYHODVRRNPLPIIARFIRVNPQWQOKTAMKMLLGCQ 404
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1378 TLFPQNG--KVKVFGNODSFTPVVNSLDPLLRLLRYLRHPQSWVHQAIALRMEVLGCE 1433
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

```

RESULT 11

```

US-10-047-257-1
; Sequence 1, Application US/10047257
; Patent No. US20020115152A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.2
; CURRENT APPLICATION NUMBER: US/10/047,257
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-047-257-1

```

```

Query Match      7.9%  Score 308.5; DB 12; Length 1438;
Best Local Similarity 37.1%; Pred. No. 2.5e-16;
Matches 66; Conservative 31; Mismatches 72; Indels 9; Gaps 3;

```

```

QY 229 SVVGHLSLSTFTKTSYGYTLGMSGVIADPQITASSVLEWTDHTGOENSKWKPKARLK 288
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1263 SIRSTLRMELMGCDLNSCSMPLGMSKASDAQITASSYF-----TNMFATWSPSKARLH 1317
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 289 KPG--PPWAAATDEYQWLQIDLNKKKITGIITGTSMVHNHYVSAYRILYSDDGOKW 346
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1318 LQGRSNAWRPQVNNPKEWLQVDFQKTMKVGTGTQGVKSLLSIMYVKEFLISSQDGHQW 1377
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 347 TVYRPGVEQDKIFOGNKDYHODVRRNPLPIIARFIRVNPQWQOKTAMKMLLGCQ 404
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 1378 TLFPQNG--KVKVFGNODSFTPVVNSLDPLLRLLRYLRHPQSWVHQAIALRMEVLGCE 1433
| : | | | : | | | : | | | : | | | : | | | : | | | : | | | :

```

RESULT 12

```

US-10-095-718-2
; Sequence 2, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22

```


7	PRIOR FILING DATE: 1998-06-24	
7	PRIOR APPLICATION NUMBER: 60/090435	
7	PRIOR FILING DATE: 1998-06-24	
7	PRIOR APPLICATION NUMBER: 60/090444	
7	PRIOR FILING DATE: 1998-06-24	
7	PRIOR APPLICATION NUMBER: 60/090445	
7	PRIOR FILING DATE: 1998-06-24	
7	PRIOR APPLICATION NUMBER: 60/090472	
7	PRIOR FILING DATE: 1998-06-24	
7	PRIOR APPLICATION NUMBER: 60/090535	
7	PRIOR FILING DATE: 1998-06-24	
7	PRIOR APPLICATION NUMBER: 60/090540	
7	PRIOR FILING DATE: 1998-06-24	
7	PRIOR APPLICATION NUMBER: 60/090542	
7	PRIOR FILING DATE: 1998-06-24	
7	PRIOR APPLICATION NUMBER: 60/090557	
7	PRIOR FILING DATE: 1998-06-24	
7	PRIOR APPLICATION NUMBER: 60/090676	
7	PRIOR FILING DATE: 1998-06-25	
7	PRIOR APPLICATION NUMBER: 60/090678	
7	PRIOR FILING DATE: 1998-06-25	
7	PRIOR APPLICATION NUMBER: 60/090690	
7	PRIOR FILING DATE: 1998-06-25	
7	PRIOR APPLICATION NUMBER: 60/090696	
7	PRIOR FILING DATE: 1998-06-25	
7	PRIOR APPLICATION NUMBER: 60/090822	
7	PRIOR FILING DATE: 1998-06-26	
7	PRIOR APPLICATION NUMBER: 60/090863	
7	PRIOR FILING DATE: 1998-06-26	
7	PRIOR APPLICATION NUMBER: 60/091360	
7	PRIOR FILING DATE: 1998-07-01	
7	PRIOR APPLICATION NUMBER: 60/091478	
7	PRIOR FILING DATE: 1998-07-02	
7	PRIOR APPLICATION NUMBER: 60/091544	
7	PRIOR FILING DATE: 1998-07-01	
7	PRIOR APPLICATION NUMBER: 60/091519	
7	PRIOR FILING DATE: 1998-07-02	
7	PRIOR APPLICATION NUMBER: 60/091626	
7	PRIOR FILING DATE: 1998-07-02	
7	PRIOR APPLICATION NUMBER: 60/091633	
7	PRIOR FILING DATE: 1998-07-02	
7	PRIOR APPLICATION NUMBER: 60/091978	
7	PRIOR FILING DATE: 1998-07-07	
7	PRIOR APPLICATION NUMBER: 60/091982	
7	PRIOR FILING DATE: 1998-07-07	
7	PRIOR APPLICATION NUMBER: 60/092182	
7	PRIOR FILING DATE: 1998-07-09	

Length 756;

Indels 27

11

DEYDCAWCACEBND

ВЫРЕПГВЕОДКИ

VKNGSG---DMI

TIPKGRPPKLTQP

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

RESULT 15
US-09-989-293A-62
; Sequence 62, Application US/09989293A
; Patent No. US2002017164A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC56
; CURRENT APPLICATION NUMBER: US/09/989,293A
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030

; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23

;; PRIOR APPLICATION NUMBER: 60/090355
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090429
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090431
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090435
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090444
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090472
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090535
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090540
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090542
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090676
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090678
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090690
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090694
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090695
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090696
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090862
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091478
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091544
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091626
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091633
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091978
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/092182
;; PRIOR FILING DATE: 1998-07-09

Query Match 5.7%; Score 220.5; DB 9; Length 756;
Best Local Similarity 31.0%; Pred. No. 1.2e-09;
Matches 58; Conservative 33; Mismatches 69; Indels 27; Gaps 7;
QY 250 LGMESGVADPOITASSVLEW--TDHTGOENSWKPKARLKKP---GPPWAFATDEYOW 304
DB 139 LGLETKITDFOLHASTVRRYGLGAHRGLNI-----QAGINENDFYDGAWCAGRNDLQOW 194
QY 305 LQIDLNKKKITGIITGTSTWYEHYVYSAYRILYSDGQKW-TVYRPPGVQDKIFQGN 363
DB 195 IEVDARRLTRFTGVITQGRNSLWSDWTSYKVMY/NSDHTWTVKNGSG---DMIFEEN 251
QY 364 KYHODVRNRLPPIIARPIRVNPTQW---QOKIAMKMLLGCQFTPKGRPPKLTQPPPPR 421
DB 252 SEKEIPVLNLPVMPVARYIRINPQSWFDNGSICMRMELTGCPL-----PDPN 299

QY 422 NSNDLKN 428
DB 300 NYHRRN 306

Search completed: January 21, 2003, 09:55:03
Job time : 15.4399 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 09:35:55 ; Search time 9.87407 Seconds
(without alignments)
8817.835 Million cell updates/sec

Title: US-10-060-830-2
Perfect score: 3969
Sequence: 1 atgcctctgttctctgct.....tttttaagaatactttga 2190

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 241982

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame_n2p.model -DRV=xlh
-Q-/cgn2_1/USPTO.spool/US10060830/runat_16012003_091241_26903/app_query.fasta_1.6933
-DB-published Applications_AA -QFMT-fastan -SUFFIX=n2p.rapp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10060830.ecgn.1.1.29.etrnat_16012003_091241_26903
-NCPU=6 -ICPU=3 -NO_XLPY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaa/CTUS_PUBCOMB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pcp.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1042	26.3	715	9	US-10-003-132-2
2	575.5	14.5	503	9	US-10-003-132-4
3	574.5	14.5	503	10	US-09-823-038A-51
4	465	11.7	458	9	US-10-003-132-6

Alignment Scores:
Pred. No.: 1.26e-81
Score: 1042.00
Percent Similarity: 52.14%
Best Local Similarity: 36.51%
Query Match: 26.25%
Length: 715
Matches: 264
Conservative: 113
Mismatches: 238
Indels: 108

ALIGNMENTS

RESULT 1
US-10-003-132-2
; Sequence 2, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-132-2

5	412	10.4	75	10	US-09-864-761-45371
6	400.5	10.1	923	12	US-10-104-440-2
7	398.5	10.0	931	12	US-10-104-440-4
8	347	8.7	2224	9	US-10-115-563-14
9	327.5	8.3	1431	12	US-10-095-718-1
10	308.5	7.8	1438	12	US-10-006-091-1
11	308.5	7.8	1438	12	US-10-047-257-1
12	308.5	7.8	1471	12	US-10-095-718-2
13	308.5	7.8	2332	9	US-09-957-641-2
14	222.5	5.6	756	9	US-09-992-598-62
15	222.5	5.6	756	9	US-09-989-293A-62
16	222.5	5.6	756	9	US-09-989-735-62
17	222.5	5.6	756	9	US-09-990-444-62
18	222.5	5.6	756	9	US-09-989-730-62
19	222.5	5.6	756	9	US-09-990-436-62
20	222.5	5.6	756	9	US-09-991-181-62
21	222.5	5.6	756	9	US-09-993-687-62
22	222.5	5.6	756	9	US-09-989-734-62
23	222.5	5.6	756	9	US-10-028-072-392
24	222.5	5.6	756	9	US-09-997-653-62
25	222.5	5.6	756	10	US-09-989-722-62
26	222.5	5.6	756	10	US-09-989-723-62
27	222.5	5.6	756	10	US-09-989-279-62
28	222.5	5.6	756	10	US-09-989-727-62
29	222.5	5.6	756	10	US-09-989-731-62
30	222.5	5.6	756	10	US-09-989-732-62
31	222.5	5.6	756	10	US-09-991-073-62
32	222.5	5.6	756	10	US-09-990-442-62
33	222.5	5.6	756	10	US-09-991-163-62
34	222.5	5.6	756	10	US-09-993-604-62
35	222.5	5.6	756	10	US-09-990-456-62
36	222.5	5.6	756	10	US-09-989-721-62
37	222.5	5.6	764	10	US-09-900-518A-2
38	205	5.2	734	9	US-10-174-590-458
39	205	5.2	734	9	US-10-176-758-458
40	205	5.2	734	12	US-10-052-586-458
41	200.5	5.1	855	10	US-09-771-161A-196
42	198	5.0	277	10	US-09-799-118-2
43	194	4.9	144	9	US-09-886-429-4
44	194	4.9	144	9	US-10-041-406-5
45	193	4.9	986	10	US-09-285-385C-19

DB: 9 Gaps: 24
US-10-060-830-2 (1-2190) x US-10-003-132-2 (1-715)
QY 13 CTCTGCTCTACTTGTCCGTCCTGCTGCTGAGAGCGTGGAGCCGACGAGCAAGTGTAT 72
DB 20 LeuAlaLeuLeuLeuAlaValSerAlaProLeuArgLeuGlnAlaGluLeuGlyAsp 39
QY 73 GGATGTGGACACACTGACTAGCCCTGAGAGTGGGAACCTTACATCCATAAATACCA 132
DB 40 GlyCysGlyHisLeuValThrTyrGlnAspSerGlyThrMetThrSerLysAsnTyrPro 59
QY 133 CAGACCTATCCCAACAGCAGCTGTTGTGAATGGAGATCGGTAAAGATGGAGAGAGA 192
DB 60 GlyThrTyrProAsnHisThrValCysGlyLysThrIleThrValProLysGlyLysArg 79
QY 193 GTTCGCATCAATTTGGTACTTTGACATTTGAAGATCTGATCTTGTGTCATTAATAC 252
DB 80 LeuIleLeuArgLeuGlyAspLeuAspIleGlu---SerGlnThrCysAlaSerAspTyr 98
QY 253 TTGAGAAATTTAATGAATGGAGTCAGCAGACACTGAAATAGCAATACTGTGCTCG 312
DB 99 LeuLeuPheThrSer-----SerSerAspGlnTyrGlyProTyrCysGly--- 113
QY 313 GGGTTGCAATGAACCATTCATTAATGAATCAAAAGGAATGAAATACATCTGCTGTCATG 372
DB 114 SerMetThrValProLysGluLeuLeuLeuLeuAsnThrSerGluValThrValArgPheGlu 133
QY 373 AGTGAATCCATCTTCTGACCGCGGATTTTGGCCCTCATPACTCTGTATPAGATAACAA 432
DB 134 SerGlySerHisIleSerGlyArgGlyPheLeuLeuThrTyrAlaSerSerAspHisPro 153
QY 433 GATCTAATTAATTTGTTGGACACTGCATCCAAATTTTGGAACTGAGTTCAGTAAGTAC 492
DB 154 AspLeuIleThrCysLeuGluArgAlaSerHisTyrLeuLysThrGluTyrSerLysPhe 173
QY 493 TGCCACGCTGGTTCGCTTCCTTTTGGTGAATATCTGGAACAATTCCTCATGATAT 552
DB 174 CysProAlaGlyCysArgAspValAlaGlyAspIleSerGlyAsnMetValAspGlyTyr 193
QY 553 AGAATTCCTCGCATCTGTCAGTGGCGTGGTGCATGAGGAGTAGTGTCAACACAGTTG 612
DB 194 ArgAspThrSerLeuLeuCysLysAlaIleAlaIleAlaGlyIleIleAlaAspGluLeu 213
QY 613 GCGCGCAATCACTGTTGTAATAGTAAAGGTATCCCTTATTAATGAACTCTTGGCT 672
DB 214 GlyGlyGlnIleSerValLeuGlnArgLysGlyIleSerArgTyrGluGlyIleLeuAla 233
QY 673 AACACCTCATCTGTTGGGACACTTATCTACAGTCTTTTACATTTAAAGACAAGT 732
DB 234 AsnGlyValLeuSerArgAspGlySerLeuSerAspLysArgPheLeuPheThrSerAsn 253
QY 733 GGATGTTATGAACACTGGGGATGGAGTCTGGTGTGATCCCGGATCTCAATAACAGCA 792
DB 254 GlyCysSerArgSerLeuSerPheGlu-----ProAspGlyGlnIleArgAla 269
QY 793 TCATCTGCTGGAGTGGAGTGCACACAGGCAAGGCAAGACAGTGTGAAACCCCAAAA 852
DB 270 SerSerSerTrpGlnSerValAsnGluSerGlyAspGlnValHisTrpSerProGlyGln 289
QY 853 GCCAGGCTGAAAACCTGGACCCCTTGGGCTGCTTTTCCCACTGATGAA----- 903
DB 290 AlaArgLeuGlnAspGlnGlyProSerTrpAlaSerGlyAspSerSerAsnHisLys 309
QY 904 ---TACAGTGGTTTACAATAGATTTCAATAAGAAAGAAAATAACAGCATTAATACC 960
DB 310 ProArgGlnTrpLeuGluIleAspLeuGlyLysLysLysIleThrGlyIleArgThr 329
QY 961 ACTGATCCCAATGGTGGACCAATTAATGCTATGCTGCTACAGAACTCTGACAGT 1020
DB 330 ThrGlySerThrGlnSerAsnPheAsnPheThrValLysSerPheValMetAsnPheLys 349
QY 1021 GATGATGGCAGAAATGGACTGTGTACAGAGAGCGCTGGTGTGGAGCAAGATAATTTT 1080

DB 350 AsnAsnAsnSerLysTrpLysThrTyrLysGlyIleValAsnAsnGluGluLysValPhe 369
QY 1081 CAAGGAAACAAGATTATCACCAGGATGCGGTAAATCACTTTTGGCCACCAATATTGCA 1140
DB 370 GlnGlyAsnSerAsnPheArgAspProValGlnAsnAsnPheIleProIleValAla 389
QY 1141 CGTTTATTAGAGTGAATCCTCCCAATGGCAGCAGAAAAATTCGCATGAAATGGAGCTG 1200
DB 390 ArgTyrValArgValValProGlnThrTyrHisGlnArgIleAlaLeuLysValGluLeu 409
QY 1201 CTCGGATGTCAGTTTATTCCTAAAGGTGCTCTCCAAACTTACTCAACCTCCACCTCCT 1260
DB 410 IleGlyCysGln-----IleThrGln----- 416
QY 1261 CGGAACAGCAATGAC-----CTCAAAAACACTACAGCCCTCCAAAATAGCC 1308
DB 417 -----GlyAsnAspSerLeuValTyrArgLysThrSerGlnSerThrSerValSerThr 434
QY 1309 AAAGTGTGTCGCCCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATGAATTTCT 1368
DB 435 LysLysGluAspGluThrIleThrArgProIle----- 445
QY 1369 GCACAGACAGACAAACAACTGCCAGTCTGATATCAGAAATACTACCGTAACCTCCAAAT 1428
DB 446 ---ProSerGluGluThr-----SerThrGlyIleAsnIleThrThr----- 458
QY 1429 GTAACCAAGATGACGGTGGCTGCGAGTTCGTCCTGTCGTCGTCATCGGTCCTCACT 1488
DB 459 -----ValAlaIleProLeuValLeuValLeuValLeuValPheAlaGlyMet 474
QY 1489 ACTCTCATCTTCATATAGTGTGCTGCTGGCACTGGAGAACAGAAAAAACAACGAA 1548
DB 475 GlyIlePheAlaAlaPhe-----ArgLysLysLysLysLysLys 486
QY 1549 GGCACCTATGACTTACTTACTGG-----GACCGGGCAGGTGGTGGAAAGGA 1596
DB 487 GlySer-----ProTyrGlySerAlaGluAlaGlnLysThrAspCysTyrLysGln 503
QY 1597 ATGAACAGTTCCTTCTGCAAAAGCAGTGACCATCAGGAAACCCCA---GTTGCGTAT 1653
DB 504 IleLysTyr-----ProPheAlaArgHis 511
QY 1654 AGCAGACGCGAA-----GTAAATCAGTCCAGTCCAAAGAGAGTCCAC-----ACA 1698
DB 512 GlnSerAlaGluPheThrIleSerTyrAspAsnGlnLysGluMetThrGlnLysLeuAsp 531
QY 1699 GTGCTGACGCTGACTCTGCAGAGTATGCTCAGCCACTGTGAGGAGAAATTTGGGTACA 1758
DB 532 LeuIleThrSerAspMetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrVal 551
QY 1759 CTTTCATCAAGATCTACCTTTAAACCA-----CAAGAAGGAAAGAGCAGGCTATGCA 1812
DB 552 ThrArgLysGlySerThrPheArgProMetAspThrAspAlaGluAlaGly---Val 570
QY 1813 GACCTAGATCCT-----TACAACCTCAGGCGGAGGAGTTTATCATGCTCTATGCT 1863
DB 571 SerThrAspAlaGlyGlyHisTyrAspCysProGlnArgAlaGlyArgHisGluTyrAla 590
QY 1864 GAACCACTCCCAATAGCGGGCTGAGTATGCAACCCCAATCATCATGACATGTCAGGG 1923
DB 591 LeuProLeuAlaProGluProGluTyrAlaThrProIleVal-----GluArg 607
QY 1924 CACCCCAACACTTCAGTTGTCAGCCCTCCACATCCACTTTCAGGCTACCGGAGCAACAA 1983
DB 608 HisValLeuArgAlaHisThrPheSerAlaGlnSerGlyTyrArgValProGlyProGln 627
QY 1984 CCT-----CCGCCACTA-----GTGGGAAT 2004
DB 628 ProGlyHisLysHisSerLeuSerSerGlyGlyPheSerProValAlaGlyValGlyAla 647
QY 2005 TACAATACACTTCTCTCCAGGACTGACAGCTGCTCTCAGCCGAGCCGAGTATGATACC 2064


```
.QY 1024 GATGGCAGAAATGGACTGTGTACAGAGAGCCTGGTGTGGAGCAAGATAAGATATTCAA 1083
      |||||
Db 244 -----ProGlyMet----- 246
.QY 1084 GGAACAAGATTATCACCAGGATGTGGTAATTAACCTTTTGGCCACCAATTATTCCAGCT 1143
      |||||
Db 246 ----- 246
.QY 1144 TTTATTAGAGTGAATCCTACCCAATGGCAGCAGAAAATTGCCATGAAAATGGAGCTGCTC 1203
      |||||
Db 246 ----- 246
.QY 1204 GGATGTCAGTTTATTCCTAAAGTGTCTCTCCAAAACCTTACTCAACCTCCACCTCCTCGG 1263
      |||||
Db 246 ----- 246
.QY 1264 AACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAATAGCAAAAGGTGTCGCCCA 1323
      |||||
Db 246 ----- 246
.QY 1324 AAATTTAGCAACCACTACAACTCGCAGTAGCAATGAATTTCTTGCACAGACAGAACAA 1383
      |||||
Db 246 ----- 246
.QY 1384 ACAACTGCCAGTCTGATATCAGAAATACCTACCGTAACCTCCAAATGTAAACCAAGATGA 1443
      |||||
Db 247 -----AsnIleThr---ThrVal 251
.QY 1444 GCGTGGCTGCAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1503
      |||||
Db 252 AlaIleProSerValIle-----PheIleAlaLeuLeuLeuThrcGlyMetGlyIlePhe 269
.QY 1504 TTAGTGTGCTTGGCACTGGAGAACAGAAAACCTGAAGCCACCTATGACTTA 1563
      |||||
Db 270 AlaIleCys-----ArgLysArgLysLysLysGlyAsn----- 280
.QY 1564 CATTAC-----TGGACCGGCGAGGTGTGGAAAGGAATGAAGCAGTTCTTCTT 1611
      |||||
Db 281 ProTyrValSerAlaAspAlaGlnLysThrGlyCysTrpLysGlnIleLysTyr----- 298
.QY 1612 CTGCAGAAAGCAGTGGACCATCAGGAACCCCA---GTTGCGTATAGCAGCAGAA--- 1665
      |||||
Db 299 -----ProPheAlaArgHisGlnSerThrGluPhe 308
.QY 1666 ---GTTAATCACTGAGTCCAAAGAGAGTCAAC-----ACAGTGTGTCAGGCTGAC 1713
      |||||
Db 309 ThrIleSerTyrAspAsnGluLysGluMetThrGlnLysLeuAspLeuIleThrSerAsp 328
.QY 1714 TCTGCAGATATGCTCAGCCACTGTGTAGGAGGAATTTGGTACACTTCATCAAGATCT 1773
      |||||
Db 329 MetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrValAlaArgLysGlySer 348
.QY 1774 ACCTTTAAACCA-----GAAGA-----GAAAGAGCAGGC 1806
      |||||
Db 349 ThrPheArgProMetAspThrAspThrGluValArgValAsnThrGluAlaSerGly 368
.QY 1807 TATGCAGACCTAGATCTTACACTCACCAGGCGAGAGTTTATCATGCTATGCTGAA 1866
      |||||
Db 369 His-----TyrAspCysProHisArgProGlyArgHisGluTyrAlaLeu 383
.QY 1867 CCACCTCCCAATACGGGGCTGAGTATGCAACCCCAATCATC----- 1908
      |||||
Db 384 ProLeuThrHisSerGluProGluTyrAlaThrProIleValGluArgHisLeuLeuArg 403
.QY 1909 -----ATGACATGTACGGGCACCCCACTTCACTTGTGTGTCAGCCCTCCACA 1956
      |||||
Db 404 AlaHisThrPheSerThrGlnSerGlyTyrArgValProGly---ProArgProThrHis 422
.QY 1957 TCCACTTTCAAGCTACGGGGACCACTCCCTCCCTAGTGGAACTTACAATACACTT 2016
      |||||
Db 423 GluHisSerHisSerGlyPheProProAlaThrGlyAlaThrGlnValGluSer 442
.QY 2017 CTCTCCAGGACTCACAGCTGCTCTCTCAGCCAGCCAGCCAGTATGATACCCCGAAGCTGGG 2076
      |||||
```

```
Db 443 TyrGlnArgProAlaSerProLysProValGlyGlyGlyTyrAspLysPro---AlaAla 461
      |||||
.QY 2077 AAGCAGAGTCTACCTGCCCCAGAGCAATTTGGTACCAGGTGCCACAGACACACAAGAA 2136
      |||||
Db 462 SerSerPheLeuAspSerArgAsp-----ProAlaSerGlnSerGln 475
.QY 2137 GTATCAGGAGCAGGAGGAGGATGGG 2160
      |||||
Db 476 MetThrSerGlyGlyAspAspGly 483

RESULT 4
US-10-003-132-6
; Sequence 6, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Shoenaker, Kimberly E.
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: NEURODILIN HOMOLOG 2CUB5
; FILE OF INVENTION: 00-62
; CURRENT APPLICATION NUMBER: US/10/003.132
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-003-132-6

Alignment Scores:
Pred. No.: 4,81e-32 Length: 458
Score: 465.00 Matches: 168
Percent Similarity: 32.49% Conservative: 75
Best Local Similarity: 22.46% Mismatches: 159
Query Match: 11.72% Indels: 346
DB: 9 Gaps: 21

US-10-060-830-2 (1-2190) x US-10-003-132-6 (1-458)
.QY 4 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 63
      |||||
Db 8 ProSerValLeuAlaLeuLeuPheAlaValCysAlaProLeuArgLeuGlnAlaGluGlu 27
.QY 64 CAAGGTGTGGTGTGGACACTGTGTAGCCCTGAGAGTGGAAACCTTACATCCATA 123
      |||||
Db 28 LeuGlyAspGlyCysGlyHisIleValThrSerGlnAspSerGlyThrMetThrSerLys 47
.QY 124 AACTACCCACAGACCTATCCAAACAGCAGTGTGTGTGAATGGAGATCCCTGTAAAGATG 183
      |||||
Db 48 AsnTyrProGlyThrTyrProAsnTyrThrValCysGluLysIleIleThrValProLys 67
.QY 184 GGAGAGAGAGTTCGCGCAATTAATTTGGTGTGACTTTGACATTTGAAGATTTCTGATTTCTTCC 243
      |||||
Db 68 GlyLysArgLeuIleLeuArgLeuGlyAspLeuAsnIleGlu---SerLysThrCysAla 86
.QY 244 TTTAATTAATTTGAGAAATTTAATGGAATTTGGATGAGCAGAACTGAAATAGCAAAATAC 303
      |||||
Db 87 SerAspTyr----- 89
.QY 304 TGTGGTCTGGGTTGCAATTAATGAACCATTAATGAATCAAAAGCAATGAATCAACATG 363
      |||||
Db 90 -----Leu 90
.QY 364 CTGTTTCATGAGTGAATCCATGTTTCTGACGCGGATTTTGGCCTCATCTCTGTATA 423
      |||||
Db 91 LeuPheSer-----SerAlaThr 96
.QY 424 GATAAACAAGATCTAAATTAATTTGTTGGACACTGCATCCAAATTTTGGAACTGAGTTC 483
      |||||
```

Db	97	AspGlnTyrAspLeuIleThrCysLeuGluArgGlySerHisTyrPheGluGluLysTyr	116
QY	484	AGTAACTACTGCCAGCTGGTTGCTGCTCTCTTTCTCGATATCTGGAACAATTCCT	543
Db	117	SerLysPheCysProAlaGlyCysArgAspIleAlaGlyAspIleSerGlyAsnThrLys	136
QY	544	CATGGATATAGAGATTCTTCGCCAATGTGCATGGCTGGTGTGCAATGCAGAGTAGTGCA	603
Db	137	AspGlyTyrArgAspThrSerLeuLeuCysLysAlaIleHisAlaGlyIleIleThr	156
QY	604	AACACGTTGGGCGGCCAATCAGTGTCTTAATTAGTAAAGGTATTCCTTATTATGAAGT	663
Db	157	AspGluLeuGlyGlyHisIleAsnLeuGlnSerLysGlyIleSerHisTyrGluGly	176
QY	664	TCCTTTGGCTAAACAAGTCACATCTCTGTGTGGACACTTATCTACAAGCTCTTTTACATTT	723
Db	177	LeuLeuAlaAsnGlyValLeuSerArgHisGlySerLeuSerGluLysArgPheLeuPhe	196
QY	724	AGACAAGTGGATGTATGGAAACACTGGGGATGGAGTCTGGTGTGATCGCGGATCCTCAA	783
Db	197	ThrThr-----	198
QY	784	ATAACAGCATCATCTGCTGGAGTGGACTGCACCACACAGGCGCAGAGACAGTGGAAA	843
Db	198	-----	198
QY	844	CCCAAAAAAGCCAGGCTGAAAAAACCTGGACCGCTTGGGCTGCTTTTGGCACTGATGAA	903
Db	198	-----	198
QY	904	TACCAGTGGTTACAATATAGATTGTAATTAAGAAAGAAATAACAGGCAATTATAACCAC	963
Db	198	-----	198
QY	964	GGATCCACCATTGGTGGAGCACAAATTACTTATGTCTGCCTACAGAAATCCTGTACAGT	1023
Db	198	-----	198
QY	1024	GATGGCGAGAAATGGACTGTGTACAGAGAGCGCTGGTGTGGAGCAAGATAAGATATTTCA	1083
Db	199	-----ProGlyMet-----	201
QY	1084	GGAACAAGATATACACAGAGATGGGTAAATTAACCTTTTGGCCACCAATATTGACGCT	1143
Db	201	-----	201
QY	1144	TTTATTAGAGTGAATCCTACCCAATGGCAGCAGAAATTCGCCATGAAATGGAGCTGCTC	1203
Db	201	-----	201
QY	1204	GGATGTCAAGTTTATTCCTAAAGGTGCTCTCCAAAACCTTACTCAACCTCCACCTCCTCGG	1263
Db	201	-----	201
QY	1264	AACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAAATAGCCAAAGTCGTGCCCCA	1323
Db	201	-----	201
QY	1324	AAATTATCCCAACCACCTACAACCTCGCAGTAGCAATGNAATTCCTGCACAGACAGACAA	1383
Db	201	-----	201
QY	1384	ACAACTGCCAGTCTGATATCAGAAATACTACCGTAACTCCCAATGTAAACCAAGATGTA	1443
Db	202	-----AsnIleThr-----ThrVal	206
QY	1444	CGCTGGCTGCAGTTCTTGTCCCTGTGTGTCATGGTCTCTCACTACTCTCAITTCATTA	1503
Db	207	AlaIleProSerValIle-----PheIleAlaLeuLeuThrGlyMetGlyIlePhe	224
QY	1504	TTAGTGTGCTTGGCACTGGGAAACAGAAAGAAAAAACTGAAGGCACCTATGACTTA	1563
Db	225	AlaIleCys-----ArgLysArgLysLysLysGlyAsn-----	235

Qy 1564 CCTTAC-----TTGGAGCCGCCGCTGGTGCGAAAGAAGNATGAACGACTTCCTT 1611
||||| :: |||| |
Db 236 ProTyValSerAlaAsPalaGlInLysThrCysTrpLysGlnLeuLysTyr----- 253
:: ||| |
Qy 1612 CTGC AAAA AG CAG TGG ACC AT GAG GA AAC CC CA --- GT TC GC TT A TG AC G AG CA AA --- 1665
||| ||||| :
Db 254 -----ProPheAlaArgHisGlnSerThrGluPhe 263
:
Qy 1666 --GTTAATCACCTGAGTCCCAAGAGAACGTACC-----ACAGTCTCTGCAGCGCTGAC 1713
:::::::::: :: ::|||:||| :
Db 264 ThrIleSerTyrAsnGluYslgslumMetThrGlnLynLeuAspLeullethrSerAsp 283
:
Qy 1714 TTCGACGAGTATGTCACCCACTGGTAGGAGAATTGTGTGTACACTTCATCAAAGATCT 1773
||||| |
Db 284 MetaIaspTyrGlnGlnProLeuMetIleGlyThrGlyThrValAlaArGLysGlySer 303
:
Qy 1774 ACCTTTAAACA-----GAAGAA-----GAAAAAGAGCAGGC 1806
||||| :
Db 304 ThrpheArgPrometAspThrAspThrGluGluValArgvalasnthrglualasergly 323
:
Qy 1807 TATGACAGACCTAGATCCTTACACTCACAGGCGAGGAAGTTTTATCATGCTCTATGCTCAA 1866
::: |||: ||| :
Db 324 His-----TyrAspcysProHIsArgProGIyArgHISglutryalaleu 338
:
Qy 1867 CCActCCAANTACGGGGCGCTGATGTCAACCCCACATC----- 1908
||||| :
Db 339 ProLeuThrHisSerGluProGluTryAlaThrProIIeValGluArgHisLeuLeuArg 358
:
Qy 1909 -----ATGCAGATGTCAGGCGACCCCCACAACTTCAGTTGGTCAGCCCTCCACA 1956
||||| :
Db 359 AlaHisThrPheSerThrGlnSerglytyrValProGly---ProArgprothrhis 377
:
Qy 1957 TCCACTTTCAGGCTACGGGAGCAACTCCCCACTAGTGGGNACTTACAATACACTT 2016
::::: ||| |
Db 378 LysHisSerHisSerSerglycylpheProAlathrGlyAlaThrGlnValGluSer 397
:
Qy 2017 CTCCTCCAGGCTGACAGCTGCTCCTCAGCCAGGCCAGTATGATACCCCGAAGCTGGG 2076
||| ||| |
Db 398 TyrGlnArgProAlaSerProlysProValGlyGlyTYraspylspPro--Alaala 416
:
Qy 2077 AAGCCAGTCTACCTGCCCCAGACGAATTGTGTACCAGGTGCCACAGAGCACACAGAA 2136
||| :: ||| :
Db 417 SerSerPheLeuAspSerArgAsp-----ProAlaSeriGlnSerGln 430
:
Qy 2137 GTAICAGGACGAGNAGGATGGG 2160
::::: ||| |
Db 431 MetThrSerglytyrAspGly 438
:
RESULT 5

US-09-864-761-45371 ; Sequence 45371, Application US/09864761
; Patent No. US20020048763A1
;
; GENERAL INFORMATION:
; APPLICANT: Penn, Shaaron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIORITY APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIORITY APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIORITY APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIORITY APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIORITY APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27

RESULT 5

US-09-864-761-45371
 ; Sequence 45371, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FC
 ; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 ; FILE REFERENCE: Aeomica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27

QY 652 -----TATTATGAAGTCTTTGGCTAACACACGTCACATCTGTG 690
Db 222 ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerGluLeuArgSer 241
QY 691 GTGGGACACTTATCTACAAAGTCTTTTACAC----- 720
Db 242 ThrGlyIleLeuSerLeuThrPheHisThrAspMetAlaValAlaLysAspGlyPheSer 261
QY 721 -----TTTAAGCAAGTGGATGTTATGTAAGAACACTG 750
Db 262 AlaArgTyrTyrLeuValHisGlnGluProLeuGluAsnPheGlnCysAsnValProLeu 281
QY 751 GGGATGGAGTCTGGTCTGATCCGGATCTCAATAACAGCATCATCTGTGGTGGAGTGG 810
Db 282 GlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer----- 297
QY 811 ACTGACCACACAGGGCAAGAACACAGTTGGAAACCCAAAGCCAGCGGTGAAAAACCT 870
Db 298 -----ThrTyrSerAspGlyArgTyrThrProGlnGlnSerArgLeuHisGlyAsp 314
QY 871 GGACCCCTTGGCTGCTTTTGGCCACTGATGATACCATCAGTGGTTACAATAGATTGTAAT 930
Db 315 AspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnValAspLeuArg 334
QY 931 AAGGAAACAAATAACAGCATATTAACCACTGGA-----TCCACCATGGTGGAGCAC 984
Db 335 PheLeuThrMetLeuThrAlaIleAlaThrGlnGlyAlaIleSerArgGluThrGlnAsn 984
QY 985 AATTACTATGTCTGCTACAGAAATCCCTGACAGATGATGGCGCAAGAAATGGACTGTG 1044
Db 355 GlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAspTrpMetVal 374
QY 1045 TACAGAGACCTGGTGGTGGAGCAAGATAAGATATTCAAGGAAACAAAGATTATCACAC 1104
Db 375 TyrArgHisGly-----LysAsnHisLysValPheGlnAlaAsnAsnAspAlaThrGlu 392
QY 1105 GATGCTCGTAATAACTTTTGGCCACCAATATTGACAGTCTTTTATGAGTGAATCCTACC 1164
Db 393 ValValLeuAsnLysLeuHisAlaProLeuLeuThrArgPheValArgIleArgProGln 412
QY 1165 CAATGCGACAGAAAATCCCATGAAATGGAGCTCTCGGATGTCAGTTTATTCCTTAA 1224
Db 413 ThrTrpHisSerGlyIleAlaLeuArgLeuGluLeuPheGlyCys----- 427
QY 1225 GTCGCTCTCCAAACTTACTCAACCTCCACCTCCGCAACAGCAATGACCTCAAAAC 1284
Db 428 -----ArgValThrAspAlaProCysSerAsnMetLeuGlyMetLeuSerGly 443
QY 1285 ACTAGACCCCTCCAAATAGCCAAAGTGTGTCGCCCAAAATTT----- 1329
Db 444 LeuIleAlaAspSerGlnIleSerAlaSerThrGlnGluTyrLeuTrpSerProSer 463
QY 1330 ACACAACCACTACACCTCGCAGTAGCAATGAAATTTCTCGCACAGACAGCAACAACT 1389
Db 464 AlaAlaArgLeuValSerSerArgSerGlyTyrPheProArgIleProGlnAlaGlnPro 483
QY 1390 GCCAGTCTCATATCAAAATACCTACCGTAACCTCCAAATGTAACCAAGATGTA----- 1443
Db 484 GlyGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGlyValIleIle 503
QY 1444 -----GGCTGGCGTCAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1482
Db 504 GlnGlyAlaArgGlyGlyAspSerIleThrAlaValAlaGluAlaArgAlaPheValArgLys 523
QY 1483 CTCACACTCTCATCTCATATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1527
Db 524 PheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAspProArgThr 543
QY 1528 AACAGAAAGAAAAAATGAGGACAC-----TATGACTTACTTCTTACTGGACCGGCA 1581
Db 544 GlnGlnProLysLeuPheGluGlyAsnMetHisTyrAspThrPro----- 558
QY 1582 GGTGTGGTGAAGGAATGAAGCAGTCTTCTTCTGCAAAAGCAGTGGACCATGAGGAAAC 1641

Db 559 -----AspIleArgArgPheAspPro-----Ile 566
QY 1642 CCAGTTTCGTATAGCAGCAGCAGGAAGTTAATCACTGATCTCAAGAGAAAGTCAACACAGT 1701
Db 567 ProAlaGlnTyrValArgValTyrProGluArgTyrPheProAlaGlyIleGlyMetArg 586
QY 1702 CTCAGGCTGACTCTGCAGAGTAT-----GCTCAGCAGCTGTTAGGAGNAATTTGTTGT 1755
Db 587 LeuGluValLeuGlyCysAspTrpThrAspSerLysProThrValGluThrLeuGlyPro 606
QY 1756 ACATCTCATCAAGATCTACTTTTAAACCA-----GAAGAAGGAAAAAGAACGA 1803
Db 607 ThrValLysSerGluGluThrThrThrProTyrProThrGluGluAlaThrGluCys 626
QY 1804 GGC-----TATGCAGACCTAGATCTCTTCACTCACTCAGGCGCAGGAGAA-- 1846
Db 627 GlyGluAsnCysSerPheGluAspLysAsp--LeuGlnLeuProSerGlyPheAsnCy 646
QY 1847 -----TTTATCATGCCT-----ATCTGAACCACTCCCAA-----TTACG 1881
Db 646 sAsnPheAspPheLeuGluProCysGlyTrpMetTyrAspHisAlaLysTrpLeuAr 666
QY 1882 GGCCTGAGTATGCAACCCCAATCATCATGACATGTGAGGCGCACCCACAA----- 1933
Db 666 g-----ThrThrTrpAlaSerSerSerSerProAsnAspArgTh 679
QY 1933 ----- 1933
Db 679 rPheProAspArgAsnPheLeuArgLeuGlnSerAspSerGlnArgGluGlyGlnTy 699
QY 1934 -CTTCACTTGGTCAGCCCTCCACATCCATCTTCAAGGCTACGGGGAACCACTCCCA 1992
Db 699 rAlaArgLeuIleSerProPro--ValHisLeu-----ProArgSerPro 714
QY 1993 CTAGTGGGAATTAACAATACACTTCTCTC----- 2021
Db 714 alCysMetGluPheGlnTyrGlnAlaThrGlyArgGlyValAlaLeuGlnValVala 734
QY 2022 -----CAGACTGACAGCTGCTC-----CTCAGCCAGCCGACCATGATATACC 2064
Db 734 rGluAlaSerGlnGluSerLysLeuLeuTrpValIle--ArgGluAspGlnGlyGlu 753
QY 2065 CCGAAAGCTGGAGCCAGGCTTACCTGCCCCCAGACGAAATTTGTTGTACCGTCCACAG 2124
Db 754 TrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGlnIle----- 770
QY 2125 AGCACAAAGATCATCAGGAGCAGGAGGATGGGAA 2163
Db 771 ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu 783
RESULT 8
US-10-115-563-14
; Sequence 14, Application US/10115563
; Publication No. US20030008307A1
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H
; Greenard, Judith S
; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATI
; AND COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; Patent Counsel
; STREET: 10666 NO. US20030008307A1th Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/115,563
; APPLICATION NUMBER: US/10/115,563
; FILING DATE: 02-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,488
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: 449.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2224 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-115-563-14

```

```

Alignment Scores:
Pred. No.: 1,27e-21 Length: 2224
Score: 347.00 Matches: 76
Percent Similarity: 58.43% Conservative: 21
Best Local Similarity: 45.78% Mismatches: 59
Query Match: 8.74% Indels: 10
DB: 9 Gaps: 3

```

US-10-060-830-2 (1-2190) x US-10-115-563-14 (1-2224)

```

QY 724 AAGACAGTGGATGTTATGAAACACTGGGATGGAGTGTGGTGTGCGGAGCTCTCAA 783
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 2062 GluValAsnGlyCysSerThrProLeuGlyMetGluAsnGlyLysLeuGluAsnLysGln 2081
QY 784 ATAAACAGCATCATCTGTGCTGGAG-----TGGACTGACACACAGCGGCAAGAACAGT 837
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 2082 IleThrAlaSerSerPheLysLysSerTrpTrpGlyAspTyr----- 2095
QY 838 TGGAAACCAAAAGCCAGCGCTGAAACACCTGGA-----CGCCTTGGGCTGCTTT 891
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 2096 TrpGluProPheArgAlaArgLeuAsnAlaGlnGlyArgValAsnAlaTrpGlnAlaLys 2115
QY 892 GCACATGATGAATACCATGCTGTACAAATAGATTGAAATAGGAAAGAAATAACAGGC 951
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 2116 AlaAsnAsnLysGlnTrpLeuGluIleAspLeuLysLysLysLysLysLysLysLysLys 2135
QY 952 ATTATACCACTGGATCCACCATGGTGGGACACAAATATATGTCTGCTACAGAAATC 1011
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 2136 IleIleThrGlnGlyCysLysSerLeuSerSerGluMetTyrValLysSerTyrThrIle 2155
QY 1012 CTCTACAGTATCATGGCAGAAATAGCTGTCTACAGAGAGCTGCTGGAGCAAGAT 1071
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 2156 HisTyrSerGlnGluValGluValGluLysProTyrArgLeuLysSerSerMetValAsp 2175
QY 1072 AAGATATTTCAAGAAACAAAGATATATCACCAGGATGCTGCTAACTTTTGGCCACCA 1131
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 2176 LysIlePheGluGlyAsnThrAsnThrLysGlyHisValLysAsnPhePheAsnProPro 2195
QY 1132 ATTATTGACGTTTATTAGAGTGAATCTTACCACCAATGGCAGCAAAATTTGCCAGAAA 1191
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 2196 IleIleSerArgPheIleArgValIleProLysThrTrpAsnGlnSerIleThrLeuArg 2215
QY 1192 ATGAGCTGCTCGGATGT 1209
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 2216 LeuGluLeuPheGlyCys 2221
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
RESULT 9
US-10-095-718-4

```

```

; Sequence 4, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

```

```

Alignment Scores:
Pred. No.: 5.11e-20 Length: 1431
Score: 327.50 Matches: 112
Percent Similarity: 41.36% Conservative: 58
Best Local Similarity: 27.25% Mismatches: 144
Query Match: 8.25% Indels: 97
DB: 12 Gaps: 15

```

US-10-060-830-2 (1-2190) x US-10-095-718-4 (1-1431)

```

QY 223 GAAGATTCATCTCTCTCACTTTAAT----- 249
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1050 GluAsnIleHisSerIleHisPheSerGlyHisValPheThrValArgLysLysGluGlu 1069
QY 250 ---TACTTGAATTTATAAT----- 267
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1070 TyrLysMetAlaValTyrAsnLeuTyrProGlyValPheGluThrValGluMetLeuPro 1089
QY 268 ---GGAATTGAGTCACGACAGACTGAA-----ATAGGCAAAATCTGTGCTGGGTTG 318
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1090 SerGlnValGlyIleTrpArgIleGluCysLeuIleGlyGluHisLeuGlnAlaGlyMet 1109
QY 319 CAAATGAACCATCAATGAATGAATCAAAAGGCAATCAATCAATTCATTCCTCATGAGTGA 378
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1110 SerThrLeuPheLeuValTyrSerLysCysGlnThrProLeuGlyMetAlaSerGly 1129
QY 379 -----ATCCATGTTTCTGGAGC----- 396
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1130 HisIleArgAspPheGlnIleThrAlaSerGlyGlnTyrGlyGlnTyrAlaProLysLeu 1149
QY 397 -----GGATTTTGGCTCATCTACTCTGTATAGAT----- 426
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1150 AlaArgLeuHisTyrSerGlySerIleAsnAlaTrpSerThrLysAspPropheSerTrp 1169
QY 427 ---AAACAAGATCTAATTAATCTGTTTG-----GACACTGCATCCAATTTTGGAA 474
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1170 IleLysValAspLeuLeuAlaProMetIleIleHisGlyIleMetThrGlnGlyAlaArg 1189
QY 475 CCTGAGTTCAGTAGTACTGCCAGCTGGTGTCTGCTCTCTCTCTGCTGAGATATCTGGA 534
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1190 GlnLysPheSerSerLeuTyrValSerGlnPheIleIleMetTyrSer---LeuAspGly 1208
QY 535 ACAATTCCTCATGATATAGAGATTCCTCGCATTTGTGCATGGTGTGTGTCATGCACGGA 594
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 1209 AsnLysTrpHisSerTyrArgGlyAsnSer----- 1218
QY 595 GTAGTGTCAAAACACGTTGGCGGCCCAATCAGTGTGTAAATAGTAAGGTATT----- 648

```

```

Db 1219 -----ThrGlyThrLeuMetValPheGlyAsnValAspSerSerGlyIleLysHis 1236
      :::: ||||| ||| ||| |||||
QY 649 -----CCCTATATGAAAGTCTTGGCTAAACAAGCTCAC-----TCT 687
      ||| ||| ||| ||| |||
Db 1237 AsnIlePheAsnProIleIleAlaGlnTyrIleArgLeuHisProThrHisTyrSer 1256
      :::: ||| ||| ||| ||| |||
QY 688 GTGGTGGGACACTTATCTACAGTCTTTTACATTTAAGACAAGTGGATGATGGAACA 747
      :::: ||| ||| ||| ||| |||
Db 1257 IleArgSerThrLeuArgMetGluLeuLeuGlyCysAspPheAsnSerCysSerMetPro 1276
      :::: ||| ||| ||| ||| |||
QY 748 CTGGGGATGGAGTCTGGTGTGATCGGATCTCAATAACAGCATCATCTGCTGGAG 807
      ||| ||| ||| ||| ||| |||
Db 1277 LeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerTyrLeu---- 1295
      :::: ||||| ||||| ||||| ||||| |||||
QY 808 TGGACTGACACACAGGGCAAGACAGTGTGAAACCCAAAAGCCAGGCTGAAAAA 867
      :::: ||||| ||||| ||||| ||||| |||||
Db 1296 -----SerSerMetLeuAlaThrTrpSerProSerGlnAlaArgLeuHisLeu 1311
      :::: ||||| ||||| ||||| ||||| |||||
QY 868 CTTGGA-----CCGCCTTGGCTGCTTTTGGCCACTGATGAATACAGTGGTTACAAATA 921
      ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1312 GlnGlyArgThrAsnAlaTrpArgProGlnAlaAsnAsnProLysGluTrpLeuGlnVal 1331
      :::: ||||| ||||| ||||| ||||| |||||
QY 922 GATTGAATAAGAAAGAAATAACAGGCATATTAACACTGGATCCACCATGGTGGAG 981
      ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1332 AspPheArgLysThrMetLysValThrGlyIleThrGlnGlyValLysSerLeuLeu 1351
      :::: ||||| ||||| ||||| ||||| |||||
QY 982 CACAATTACTATGTCTGCCCTACAGATCCCTGTACAGTGTGATGGCGAGAATGGACT 1041
      :::: ||||| ||||| ||||| ||||| |||||
Db 1352 IleSerMetTyrValLysGluPheLeuIleSerSerGlnAspGlyHisAsnTrpThr 1371
      :::: ||||| ||||| ||||| ||||| |||||
QY 1042 GTGTACAGAGCCTGGTGTGGAGCAAGATAAGATATTCAAGGAAACAAGATTATCAC 1101
      :::: ||||| ||||| ||||| ||||| |||||
Db 1372 LeuPheLeuGlnAsnGly-----LysValLysValPheGlnGlyAsnArgAspSer 1389
      :::: ||||| ||||| ||||| ||||| |||||
QY 1102 CAGGATGTCGTAATTAACATTTTGGCCACCAATATTGACAGTTTATTAGATGATCCT 1161
      ||||| ||||| ||||| ||||| ||||| |||||
Db 1390 ThrProValArgAsnArgLeuGluProProLeuValAlaArgTyrValArgLeuHisPro 1409
      :::: ||||| ||||| ||||| ||||| |||||
QY 1162 ACCAATGTCGACAGAAATGTCATGAAATGCGATGCTCGGATGTCAGTTTATTTCCT 1221
      ||||| ||||| ||||| ||||| ||||| |||||
Db 1410 GlnSerTrpAlaHisIleAlaLeuArgLeuGluValLeuGlyCysasp----- 1426
      :::: ||||| ||||| ||||| ||||| |||||
QY 1222 AAAGTGTCTCTCCAAACTTACTCAACCTCCA 1254
      ||||| |||
Db 1427 -----ThrGlnGlnPro 1430
      :::: ||||| |||

```

RESULT 10

US-10-006-091-1

; Sequence 1, Application US/10006091

; Patent No. US20020102730A1

; GENERAL INFORMATION:

; APPLICANT: Cho, Myung-Sam

; APPLICANT: Chan, Sham-Yuen

; APPLICANT: Kelsey, William

; APPLICANT: Yee, Helena

; TITLE OF INVENTION: Expression System for Factor VIII

; FILE REFERENCE: MSB-7255.1

; CURRENT APPLICATION NUMBER: US/10/006,091

; CURRENT FILING DATE: 2001-12-06

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1438

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Derived from

; OTHER INFORMATION: human factor VIII sequence

US-10-006-091-1

Alignment Scores:

Pred. No.:

2.21e-18

Length:

1438

Score: 308.50 Matches: 66
 Percent Similarity: 54.49% Conservative: 31
 Best Local Similarity: 37.08% Mismatches: 72
 Query Match: 7.77% Indels: 9
 DB: 12 Gaps: 3

US-10-060-830-2 (1-2190) x US-10-006-091-1 (1-1438)

QY 685 TCTGTGGGGACACTTATCTCAAGTCTTTTACATTTTAAAGACAAGTGTGATGTTATGGA 744

Db 1263 SerIleArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet 1282

QY 745 ACACATGGGGATGGAGTCTGGTGTGATCGGATCCCTCAATAACAGCATCATCTGCTGCTG 804

Db 1283 ProLeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerTyrPhe 1302

QY 805 GAGTGGACTGACACACAGGGCAAGACAGTGTGAAACCCAAAAGCCAGGCTGAA 864

Db 1303 -----ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis 1317

QY 865 AAACCTGGA-----CCGCCTTGGGCTGCTTTTGGCCACTGATGAATACACAGTGGTTACAA 918

Db 1318 LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGln 1337

QY 919 ATAGATTTCAATTAAGAAAGAAATAACAGGCATATTAACACTGGATCCACCATGGTG 978

Db 1338 ValAspPheGlnLysThrMetLysValThrGlyValThrGlnGlyValLysSerLeu 1357

QY 979 GAGCACAATTACTATGTCTGCTGCATCAGAAATCTGTACAGTGTGATGGCGAGAATGG 1038

Db 1358 LeuThrSerMetTyrValLysGluPheLeuIleSerSerGlnAspGlyHisGlnTrp 1377

QY 1039 ACTGTGTACAGAGAGCCTGGTGTGGAGCAAGATAAGATATTCAAGGAAACAAGATTAT 1098

Db 1378 ThrLeuPhePheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer 1395

QY 1099 CACCAGATGTGCGTAAATAACTTTTGGCCACCAATATTGACAGTTTATTAGATGCAAT 1158

Db 1396 PheThrProValValAsnSerLeuAspProProLeuLeuThrArgTyrLeuArgIleHis 1415

QY 1159 CPTACCAATGGCAGCAAGAAATGGCCATGAAATGGAGTGTGCGGATGTCAG 1212

Db 1416 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 1433

RESULT 11

US-10-047-257-1

; Sequence 1, Application US/10047257

; Patent No. US20020115152A1

; GENERAL INFORMATION:

; APPLICANT: Cho, Myung-Sam

; APPLICANT: Chan, Sham-Yuen

; APPLICANT: Kelsey, William

; APPLICANT: Yee, Helena

; TITLE OF INVENTION: Expression System for Factor VIII

; FILE REFERENCE: MSB-7255.2

; CURRENT APPLICATION NUMBER: US/10/047,257

; CURRENT FILING DATE: 2002-01-15

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1438

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Derived from

; OTHER INFORMATION: human factor VIII sequence

US-10-047-257-1

Alignment Scores:

Pred. No.:

2.21e-18

Length:

1438

Score:

308.50

Matches:

66

Conservative:

31

Mismatches:

72

Alignment Scores:

Pred. No.:

2.21e-18

Length:

1438

Query Match: 7.77% Indels: 9
DB: 12 Gaps: 3

US-10-060-830-2 (1-2190) x US-10-047-257-1 (1-1438)

QY 685 TCTGTGGTGGGACACTTATCTACAGTCTTTTACATTTTAAGACAAGTGGATGTTATGGA 744
DB 1263 SerIleArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet 1282
QY 745 ACACCTGGGATGAGTCTGGTGTGATCGCGGATCCTCAATAACAGCATCATCTGTGCTG 804
DB 1283 ProLeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyrPhe 1302
QY 805 GAGTGGACTGACACACAGGCGCAAGACAGTGTGAACCCCAAAAGCCAGGCTGAAA 864
DB 1303 -----ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis 1317
QY 865 AAACCTGGGA-----CGGCTTGGGCTGCTTTTCCCACTGATGAATACCAGTGGTTACAA 918
DB 1318 LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGln 1337
QY 919 ATAGATTGAATGAAGAAAGAAATAACAGGCATTATACCACTGATCCACCATGCTG 978
DB 1338 ValAspPheGlnLysThrMetLysValThrGlyValThrThrGlnGlyValLysSerLeu 1357
QY 979 GAGCACAATTACTATGCTGCTGCTACAGAACTCTGTACAGTATGATGGCGCAAAATGG 1038
DB 1358 LeuThrSerMetTyrValLysGluPheLeuIleSerSerGlnAspGlyHisGlnTrp 1377
QY 1039 ACTGTGTACAGAGAGCTGCTGTGGAGCAAGATATTTCAAGGAACAAGATTTAT 1098
DB 1378 ThrLeuPheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer 1395
QY 1099 CACAGAGATGCGTATAACTTTTTCGCCACCAATATTGCGAGTGTATTTAGAGTGAAT 1158
DB 1396 PheThrProValValAsnSerLeuAspProProLeuLeuThrArgTyrLeuArgIleHis 1415
QY 1159 CCTACCAATGGCAGCAGCAAAATGCCATGAATGAGCTGCTCGATGCTCAG 1212
DB 1416 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 1433

RESULT 12

US-10-095-718-2
; Sequence 2, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1471
; TYPE: PRT
; ORGANISM: Homo sapiens B-domain deleted factor VIII
; FEATURE:
; OTHER INFORMATION: Homo sapiens BDD FVIII
US-10-095-718-2
Alignment Scores: 2.23e-18 Length: 1471
Pred. No.: 2.23e-18 Length: 1471
Score: 308.50
Percent Similarity: 54.49%
Best Local Similarity: 37.08%
Query Match: 7.77%
DB: 9

Score: 308.50 Matches: 66
Percent Similarity: 54.49% Conservative: 31
Best Local Similarity: 37.08% Mismatches: 72
Query Match: 7.77% Indels: 9
DB: 12 Gaps: 3

US-10-060-830-2 (1-2190) x US-10-095-718-2 (1-1471)

QY 685 TCTGTGGTGGGACACTTATCTACAGTCTTTTACATTTTAAGACAAGTGGATGTTATGGA 744
DB 1296 SerIleArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet 1315
QY 745 ACACCTGGGATGAGTCTGGTGTGATCGCGGATCCTCAATAACAGCATCATCTGTGCTG 804
DB 1316 ProLeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyrPhe 1335
QY 805 GAGTGGACTGACACACAGGCGCAAGACAGTGTGAACCCCAAAAGCCAGGCTGAAA 864
DB 1336 -----ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis 1350
QY 865 AAACCTGGGA-----CGGCTTGGGCTGCTTTTCCCACTGATGAATACCAGTGGTTACAA 918
DB 1351 LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGln 1370
QY 919 ATAGATTGAATGAAGAAAGAAATAACAGGCATTATACCACTGATCCACCATGCTG 978
DB 1371 ValAspPheGlnLysThrMetLysValThrGlyValThrThrGlnGlyValLysSerLeu 1390
QY 979 GAGCACAATTACTATGCTGCTGCTACAGAACTCTGTACAGTATGATGGCGCAAAATGG 1038
DB 1391 LeuThrSerMetTyrValLysGluPheLeuIleSerSerGlnAspGlyHisGlnTrp 1410
QY 1039 ACTGTGTACAGAGAGCTGCTGTGGAGCAAGATATTTCAAGGAACAAGATTTAT 1098
DB 1411 ThrLeuPheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer 1428
QY 1099 CACAGAGATGCGTATAACTTTTTCGCCACCAATATTGCGAGTGTATTTAGAGTGAAT 1158
DB 1429 PheThrProValValAsnSerLeuAspProProLeuLeuThrArgTyrLeuArgIleHis 1448
QY 1159 CCTACCAATGGCAGCAGCAAAATGCCATGAATGAGCTGCTCGATGCTCAG 1212
DB 1449 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 1466

RESULT 13

US-09-957-641-2
; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
; CURRENT FILING DATE: 2001-09-16
; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-957-641-2
Alignment Scores: 2.67e-18 Length: 2332
Pred. No.: 2.67e-18 Length: 2332
Score: 308.50
Percent Similarity: 54.49%
Best Local Similarity: 37.08%
Query Match: 7.77%
DB: 9

[illegible]

QY 1312 GGTCGTCCCAAAATTTAGCGAACCACTACAACTCGCAGTAGCAATGAATTCCTGCA 1371
::
Db 316 -----PhelyshHisasn-----Ty 321
::
QY 1372 CAGACAGACAACAACACTCGCATCTGTATATCAGAAATACTACCG 1417
:::
Db 321 rLysGluMetArgGlnLeuMetLysValValAsnGluMetCysPro 336
:::
RESULT 15
US-09-989-293A-62
; Sequence 62, Application US/09989293A
; Patent No. US20020177164A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC66
; CURRENT APPLICATION NUMBER: US/09/989,293A
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 09:35:55 ; Search time 10.2799 Seconds
(without alignments)
8817.835 Million cell updates/sec

Title: US-10-060-830-1
Perfect score: 4143
Sequence: 1 gccgcgcgcgcgcgcctggc.....tttgaagatgatgctgcttt 2280

Scoring table:	BLOSUM62				
	Xgapop	10.0	Xgapext	0.5	
	Ygapop	10.0	Ygapext	0.5	
	Egapop	6.0	Egapext	7.0	
	Delop	6.0	Delext	7.0	

Searched: 120991 seqs, 19878514 residues
Total number of hits satisfying chosen parameters: 241982

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```

Command line parameters:
-MODEL=framet_n2p.model -DEV=xlh
-O/-cgn2_1/USPTO.spool/US10060830/runat_16012003_091241_26903/app_query.fasta_1.6933
-DB=published.Applications_AA -SUFFIX=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloss62
-TRANS=human40 cdi_115T=45 -FQCALIGN=200 -THR_SCORE=pcet -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10060830 -ECGN_11_29 -runat_16012003_091241_26903
-NCPU=6 -ICPU=3 -NO_XLPHY -LARGESQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

```

```

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	1042	25.2	715	9 US-10-003-132-2	Sequence 2, Appl
2	575.5	13.9	503	9 US-10-003-132-4	Sequence 4, Appl
3	574.5	13.9	503	10 US-09-823-038A-51	Sequence 51, Appl
4	465	11.2	458	9 US-10-003-132-6	Sequence 6, Appl

5	412	9-9	75	10	US-09-864-761-45371	Sequence 45371, A
6	401.5	9-7	923	12	US-10-104-440-2	Sequence 2, Appl
7	398.5	9-6	931	12	US-10-104-440-4	Sequence 4, Appl
8	347	8-4	2224	9	US-10-115-563-14	Sequence 14, Appl
9	327.5	7-9	1431	12	US-10-095-718-4	Sequence 4, Appl
10	308.5	7-4	1438	12	US-10-006-091-1	Sequence 1, Appl
11	308.5	7-4	1438	12	US-10-047-257-1	Sequence 1, Appl
12	308.5	7-4	1471	12	US-10-095-718-2	Sequence 2, Appl
13	308.5	7-4	2332	9	US-09-957-641-2	Sequence 2, Appl
14	222.5	5-4	756	9	US-09-992-598-62	Sequence 62, Appl
15	222.5	5-4	756	9	US-09-989-293A-62	Sequence 62, Appl
16	222.5	5-4	756	9	US-09-989-735-62	Sequence 62, Appl
17	222.5	5-4	756	9	US-09-980-444-62	Sequence 62, Appl
18	222.5	5-4	756	9	US-09-989-730-62	Sequence 62, Appl
19	222.5	5-4	756	9	US-09-990-436-62	Sequence 62, Appl
20	222.5	5-4	756	9	US-09-991-181-62	Sequence 62, Appl
21	222.5	5-4	756	9	US-09-993-687-62	Sequence 62, Appl
22	222.5	5-4	756	9	US-09-989-734-62	Sequence 62, Appl
23	222.5	5-4	756	9	US-10-028-072-392	Sequence 392, Appl
24	222.5	5-4	756	9	US-09-987-653-62	Sequence 62, Appl
25	222.5	5-4	756	10	US-09-989-722-62	Sequence 62, Appl
26	222.5	5-4	756	10	US-09-989-723-62	Sequence 62, Appl
27	222.5	5-4	756	10	US-09-989-279-62	Sequence 62, Appl
28	222.5	5-4	756	10	US-09-989-727-62	Sequence 62, Appl
29	222.5	5-4	756	10	US-09-989-731-62	Sequence 62, Appl
30	222.5	5-4	756	10	US-09-989-732-62	Sequence 62, Appl
31	222.5	5-4	756	10	US-09-991-073-62	Sequence 62, Appl
32	222.5	5-4	756	10	US-09-990-442-62	Sequence 62, Appl
33	222.5	5-4	756	10	US-09-991-163-62	Sequence 62, Appl
34	222.5	5-4	756	10	US-09-993-604-62	Sequence 62, Appl
35	222.5	5-4	756	10	US-09-990-456-62	Sequence 62, Appl
36	222.5	5-4	756	10	US-09-989-721-62	Sequence 62, Appl
37	222.5	5-4	764	10	US-09-900-518A-2	Sequence 2, Appl
38	205	4-9	734	9	US-10-174-590-458	Sequence 458, Appl
39	205	4-9	734	9	US-10-176-758-458	Sequence 458, Appl
40	205	4-9	734	12	US-10-052-586-458	Sequence 458, Appl
41	200.5	4-8	855	10	US-09-771-161A-196	Sequence 196, Appl
42	198	4-8	277	10	US-09-799-118-2	Sequence 2, Appl
43	195	4-7	144	9	US-09-919-497-89	Sequence 89, Appl
44	194	4-7	149	4	US-09-886-429-4	Sequence 4, Appl
45	194	4-7	144	9	US-10-041-406-5	Sequence 5, Appl

ALIGNMENTS

```

RESULTS
US-10-003-132-2
; Sequence 2, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG 2CUBS
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-132-2

```

Alignment Scores:		
Pred. No.:	2,72e-77	Length: 715
Score:	1042.00	Matches: 264
Percent Similarity:	52.14%	Conservative: 113
Best Local Similarity:	36.51%	Mismatches: 238
Query Match:	25.13%	Indels: 108

	9	24
DB:	Gaps:	
<hr/>		
US-10-060-830-1 (1-2280) x US-10-003-132-2 (1-715)		
Qy	88 CTCTGCTCTTACTTGCTGTCTGCTGTCGCAGGACGCTGGAGCCCGCAGCAAGGTGAT	147 : :::
Db	20 LeuAlaLeuLeuLeuLaValSerAlaProLeuArgLeuGlnAlaGluLuLeuGlyAsp	39 : :::
Qy	148 GGATGTGGACACACTGTACTAGGCCCTGAGAGTAGGAACCCCTTACCATTACAATAACCCA	207 : :::
Db	40 GlyCysGlyHisLeuValThrTyrgInAspSergIleThrMetThrSerLysAsnTyPPro	59 : :::
Qy	208 CAGACTATCCC AACAGCACACTGTTGTGAATGGGAGATCCGTGPAAGATGGGAGAGA	267 : :::
Db	60 GlyThrTyPProAsnHisThrValCysGluLysThrIleThrValProLysGlyLysArg	79 : :::
Qy	268 GTTCGCATCAAATTCGTGCTGACCTTGACATGGAAGATTCGTCTGTCACCTTTAAATTA	327 : :::
Db	80 LeulleuArgLeuGlyAspLeuspileGlu---SergInlthrCysAlaSerAspTyR	98 : :::
Qy	328 TTGAGAATTTAATAGGAATTGGAGTCACGAGAAGCTGAATAGGCAAACTACTGTGTCG	387 : :::
Db	99 LeuLeupheThrSer-----SerSerAspGlnTyRGlyProTyRCysGly---	113 : :::
Qy	388 GGGTGC AATGAACCATTC AATTGAATGA CAAMAGCAAT GAATCA CATGCTGCTTCATG	447 : :::
Db	114 SerMetThrValProLysGluLeuLeuLeuLeuAsnThrSergIuValThrValArgpheGlu	133 : :::
Qy	448 AGTGGAAATCCATCTTCTGGACGCGATTTTTGGCTCATACTCTGTTATAGATAAACAA	507 : :::
Db	134 SerySerHisIleserGlyArgGlyPheLeuLeuthrTyRAlaSerSerAspHispro	153 : :::
Qy	508 GATCAATTAATCTGTTGGACACTGCATCCAAATTTTTGGAAACCTGAGTTTCAGTAAGTAC	567 : :::
Db	154 AspLeuillehrCysLeuGuargAlaSerHisTyRLeuLysThrGluTyRSerLysPhe	173 : :::
Qy	568 TGCCGACGCTGGTGTCTGCTTCCCTTTGCTGAGATATCTGGAACAAATTCCTCATPGGNAT	627 : :::
Db	174 CysProalaglyCysArgaspValalaglialglialserGlyAsnMetValaspglytyr	193 : :::
Qy	628 AGAATTCFCGCCANTGTGCATGGCTGGTGGCATGCAGAGTAGTCTCAACACCTG	687 : :::
Db	194 ArgaspThrSerLeuleucysylsAlaIleHisAlaGlylleleAlaaspgluLeu	213 : :::
Qy	688 GCGCGCAAATCAGTGTGTAAATAGTAAGTATTCCTATTAGAAAGTTCCTTGCGCT	747 : :::
Db	214 GlyglyGlnileSerValleuGlnArgLysGlylleSerArgTyrglyGlylleLeuala	233 : :::
Qy	748 AACACGCTCATCTCTGTGTGGACACTTATCTACAAGCTCTTTTACATTTTAAGACAAGT	807 : :::
Db	234 AsnglyVallLeuSerArgaspGlySerLeuSerAspLysArgPheLeupheThrSerasn	253 : :::
Qy	808 GGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATCGCGGATCCTCAATTAACAGCA	867 : :::
Db	254 GlyCysSerArgSerLeuSerPheGlu-----ProaspglyGlnIleArgAla	269 : :::
Qy	868 TCATCTGCTGAGTGGAGTACACACAGGCGCAAGACAGTTGGAAAACCAAAAAA	927 : :::
Db	270 SerSerSerTrpGlnSerValansGluSerGlyAspGlnValHistRpSerProGlyGln	289 : :::
Qy	928 GCCAGGTGAAAAA ACTGGACCGCTTGGCGTCTGTTTGGCCACTGTAGAA-----	978 : :::
Db	290 AlaArgLeuGlnAspGlnGlyProSerTrpAlaSerGlyAspSerSerAsnAsnHisllys	309 : :::
Qy	979 ---TACACAGTGGTTACATAGATTTGAATAGGAAAAGAAATAACAGGCAATTATACC	1035 : :::
Db	310 ProArgGlnTrpLeuGlnIleAspLeuGlyLuLysLysIleThrGlyIleArgThr	329 : :::
Qy	1036 ACTGGATCCACCATGGTGGGACACATTA TCTATGCTGCTG CCTACAGAATCCTGTACAGT	1095 : :::
Db	330 ThrGlySerThrGlnSerAsnPheAsnPheTyRValLysSerPheValMeCasnpPheLys	349 : :::
Qy	1096 GATGATGGCGAGAAATGGACTGTGTACAGAGAGCCTGGTGTGGAGCAAGATAAGATATT	1155 : :::

Db	350	AsnAsnAsnSerLysTrpLysThrTyrlsGlylleValAsnGluGluLysValPhe	369
Qy	1156	CAAGGAAACAAGAATGATCACCAGGATGCGCGTAATAACATTTTGGCACCAATATTTGCA	1215
Db	370	GlnGlyAsnSerAsnPheArgPspProValGlnAsnAsnPheIleProIleValAla	389
Qy	1216	CGTTTTATTAGAGTGAATCCTACCNAATGCGAGCAGAAATTCGCATGAAATGGAGCTG	1275
Db	390	ArgTyrlValArgValProGlnThrTrpHisGlnArgIleAlaLeuLysValGluLeu	409
Qy	1276	CTCGAGTCGAGTTTATCTTAAAGTGTCTCTCCAAACCTTACTCAACCTCCACCTCCT	1335
Db	410	IleGlyCysGln-----IleThrGln-----	416
Qy	1336	CGGACACGCAATGAC-----CTCAAAACACTACAGCCCTCCAAAAATAGCC	1383
Db	417	-----GlyAsnAspSerLeuValTrpArgLysThrSerGlnSerThrSerValSerThr	434
Qy	1384	AAAGTCGTGCCCCAAAATTTACGCAACCACTACAACTCGCAGTAGCAATGAAATTCCT	1443
Db	435	LysLysGluAspGluThrIleThrArgProIle-----	445
Qy	1444	GCACAGACAGACAACAACACTGCCAGTCCTGATATCATCAGAAATACTACCGTAACCTCCAAT	1503
Db	446	---ProSerGluGluThr-----SerThrGlyIleAsnIleThr-----	458
Qy	1504	GTAACCAAGATGATGCGCTGCGTCACTTCTTGTCCTCGTGTGGTCATGCTCCTCACT	1563
Db	459	-----ValAlaIleProLeuValLeuValValLeuValPheAlaGlyMet	474
Qy	1564	ACTCTCATTCATATAGTGTGTGCTGCGACTGGAGAACAGAAAGAAAAAAGCTGAA	1623
Db	475	GlyIlePheAlaAlaPhe-----ArgLysLysLysLysLys	486
Qy	1624	GGCACTATGACTTACCTTACTTGG-----GACCGGCGAGCTTGGTGGAAAGGA	1671
Db	487	GlySer-----ProTyrlGlySerAlaGluAlaGlnLysThrAspCysTrpLysGln	503
Qy	1672	ATGAAAGCAGTTTCTCTCGAAAAAGCAGTGGACCATGAGGAAACCCCA---GTTTCGGTAT	1728
Db	504	IleLysTyr-----ProPheAlaArgHis	511
Qy	1729	AGCAGCAGCGAA---GTTATCATCTGAGTGCACAGAGAAGTCAC---ACA	1773
Db	512	GlnSerAlaGluPheThrIleSerTyrlAspAsnGluLysGluMetThrGlnLysLeuAsp	531
Qy	1774	GTGCTGCAGGCTGACTCTCGAGATGCTCAGCCACTGTTAGGAGGAATTTGGTGATCA	1833
Db	532	LeuIleThrSerAspMetAlaAspTyrlGlnGlnProLeuMetIleGlyThrGlyThrVal	551
Qy	1834	CTTCATCAAGATCTACCTTTTAAACCA-----GAAGAAGGAAAGACGAGGCTATGCA	1887
Db	552	ThrArgLysGlySerThrPheArgPrometAspThrAspAlaGluAlaGly---Val	570
Qy	1888	GACCTAGATCCT-----TACAACCTACCAAGCGGAGGAATTTATCATGCTATGCT	1938
Db	571	SerThrAspAlaGlyGlyHisTyrlAspCysProGlnArgAlaGlyArgHisGluTyrlAla	590
Qy	1939	GAACCACTCCCAATTCAGGGGCTGATGTGCAACCCCAATCATCATGCGACATGTCAGGG	1998
Db	591	LeuProLeuAlaProProGluProGluTyrlAlaThrProIleVal-----GluArg	607
Qy	1999	CACCCCAACAACCTCAGTTGGTCAGCCCTCCACATCCACTTTCGAAGGCTACGGGCAACCA	2058
Db	608	HisValLeuArgAlaHisThrPheSerAlaGlnSerGlyTyrlArgValProGlyProGln	627
Qy	2059	CCT-----CCCCCACTA-----GTGGGAACCT	2079
Db	628	ProGlyHisLysHisSerLeuSerSerGlyGlyPheSerProValAlaGlyValGlyAla	647
Qy	2080	TACAATACACTTCTCTCCAGGACTGACAGCTGCTCCTCAGCGGCCCTAGTATGTATACC	2139


```
QY 1099 GATGGCAGAAATGACTGTGTACAGAGAGCCTGGTGTGGAGCAAGATAAGATATTCAA 1158
Db 244 -----ProGlyMet----- 246
QY 1159 GGAACAAGATTATCACCAGGATGTGGTAATAACTTTTCCACCAATTATTGCACGT 1218
Db 246 ----- 246
QY 1219 TTTATTAGAGTGAATCCTACCAATGGCAGCAGAGAAATGGCCATGAAATGGAGCTGCTC 1278
Db 246 ----- 246
QY 1279 GGATGTCAGTTTATTCCTAAAGGTGCTCTCCAAACTTACTCAACCTCCACCTCTCGG 1338
Db 246 ----- 246
QY 1339 AACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAATAGCCAAAGGTCTGTGCCCA 1398
Db 246 ----- 246
QY 1399 AAATTTAGCAACCACTACAACTCGCAGTAGCAATGAATTCCTGCACAGACAGAACAA 1458
Db 246 ----- 246
QY 1459 ACAACTGCCAGTCTGATATCAGAAATACTACCGTAACCTCAAAATGTAACCAAGATGTA 1518
Db 247 -----AsnIleThr---ThrVal 251
QY 1519 GCGCTGGCTGAGTCTTCTCCTGTGTGTCATGCTGCTCCTCACTACTCTCATTTCTCATA 1578
Db 252 AlaIleProSerValIle-----PheIleAlaLeuLeuLeuThrGlyMetGlyIlePhe 269
QY 1579 TTAGTGTGTGCTTGGCACTGGAGAAACAGAAAGAAAGAACTGAAGCCACCTATGACTTA 1638
Db 270 AlaIleCys-----ArgLysArgLysLysLysGlyAsn----- 280
QY 1639 CTTTAC-----TGGGACGGGCGAGTGTGGTGAAGGAATGAAGCAGTTTCTT 1686
Db 281 ProTyrValSerAlaAspAlaGlnLysThrGlyCysTrpLysGlnIleLysTyr----- 298
QY 1687 CTGCAAAAGCAGTGGACCATGAGNAACCCCA---GTTGCGTATAGCAGCAGCGAA--- 1740
Db 299 -----ProPheAlaArgHisGlnSerThrGluPhe 308
QY 1741 ---GTTAATCACTGAGTCCAAAGAGTCAAC-----ACAGTGTGTCAGGCTGAC 1788
Db 309 ThrIleSerTyrAspAsnGlnLysGluMetThrGlnLysLeuAspLeuIleThrSerAsp 328
QY 1789 TGTGACAGATGTCTCAGCCACTGTGTAGGAGGAATTTGGTACACTTTCATCAAGATCT 1848
Db 329 MetAlaAspTyrGlnGlnProLeuMetIleGlyThrGlyThrValAlaArgLysGlySer 348
QY 1849 ACCTTTAAACCA-----GAAGA-----GGAAGAAGACGAGGC 1881
Db 349 ThrPheArgProMetAspThrAspThrGluGluValArgValAsnThrGluAlaSerGly 368
QY 1882 TATGCAGACCTAGATCCTTACAACTCACCAGGCGAGGAGTTTATCATGCTATGCTGAA 1941
Db 369 His-----TyrAspCysProHisArgProGlyArgHisGluTyrAlaLeu 383
QY 1942 CCACCTCCCAATTACGGGGCTGAGTATGCAACCCCAATCATC----- 1983
Db 384 ProLeuThrHisSerGluProGluTyrAlaThrProIleValGluArgHisLeuLeuArg 403
QY 1984 -----ATGGACATGTCAGGCGACCCCACTTCAGTTTGGTCCAGCCCTCCACA 2031
Db 404 AlaHisThrPheSerThrGlnSerGlyTyrArgValProGly---ProArgProThrHis 422
QY 2032 TCCACTTTTCAGGCTACGGGACCAACCTCCCCACTAGTGGGAACTTACATACACTT 2091
Db 423 GluHisSerHisSerSerGlyGlyPheProProAlaThrGlyAlaThrGlnValGluSer 442
QY 2092 CTCTCCAGGACTGACAGCTGCTCTCAGCCCGAGCCCGCATGATACCCCGAAAGCTGG 2151
```

```
Db 443 TyrGlnArgProAlaSerProLysProValGlyGlyTyrAspLysPro---AlaAla 461
QY 2152 AAGCCAGGCTACTCTGCCAGAGCAATGGTGTACAGGTGCCAGACAGACACAAGAA 2211
Db 462 SerSerPheLeuAspSerArgAsp-----ProAlaSerGlnSerGln 475
QY 2212 GTATCAGGACGAGGAAGGATGGG 2235
Db 476 MetThrSerGlyGlyAspAspGly 483
RESULT 4
US-10-003-132-6
; Sequence 6, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-003-132-6
Alignment Scores:
Pred. No.: 3 54e-30 Length: 458
Score: 465.00 Matches: 168
Percent Similarity: 32.49% Conservative: 75
Best Local Similarity: 22.46% Mismatches: 159
Query Match: 11.22% Indels: 346
DB: 9 Gaps: 21
US-10-060-830-1 (1-2280) x US-10-003-132-6 (1-458)
QY 79 CCTCTGTCTCCTGCTCTTACTGTCTCTGCTCGAGGAGCGCTGGAGCCGAG 138
Db 8 ProSerValLeuAlaLeuLeuPheAlaValCysAlaProLeuArgLeuGlnAlaGlu 27
QY 139 CAAGGTGATGGATGTGCACACACTGACTAGCGCCTGAGAGTGAACCTTACATCATA 198
Db 28 LeuGlyAspGlyCysGlyHisIleValThrSerGlnAspSerGlyThrMetThrSerLys 47
QY 199 AACTTACCCACAGACTATCCCAACAGCAGCTGTTGTGAATGGGAGATCCCTGTAAGATG 258
Db 48 AsnTyrProGlyThrTyrProAsnTyrThrValCysGluLysIleIleThrValProLys 67
QY 259 GGAGAGAGTTCGCATCAAAATTTGGTACTTTGCATTTGAAGATTTGATTTCTTGCAC 318
Db 68 GlyLysArgLeuIleLeuArgLeuGlyAspLeuAsnIleGlu---SerLysThrCysAla 86
QY 319 TTTAATTAATTTGAGATTTTATATGGAATTTGGATGAGTACAGAGAACTGAAATAGCAATAC 378
Db 87 SerAspTyr----- 89
QY 379 TGTGCTGGGTGGTGGCAATCAATGAATCAATGAATCAAAAGCAATGAATCAATG 438
Db 90 -----Leu 90
QY 439 CTGTTTCATGAGTGAATCCATGTTTCTGAGCGGGATTTTGGCCTCATCTCTGTATA 498
Db 91 LeuPheSer-----SerAlaThr 96
QY 499 GATAAACAAGATCTAATTACTTGTGTTGGACACTGCATCCCAATTTTGGAACTGAGTTC 558
Db 499 GATAAACAAGATCTAATTACTTGTGTTGGACACTGCATCCCAATTTTGGAACTGAGTTC 558
```

Qy	1639	CTTAC-----TGGGACCGGCGAGTGGTGCAAAAGGAATGACGAGCTTTC	1686
Db	236	ProtyrValSerAlaAspAlaGlnLysThrGlyCysTrpLysGlnIleLysyr-----	253
Qy	1687	CCTCAAAAGCAGTGGACCATGAGAAACCCCA--GTTGCGTATAGCAGCAGCAA---	1740
Db	254	-----ProPheAlaArgHisGlnSerThrGlnPhe	263
Qy	1741	---GTTAATCACCTGAGTCCACAGAGAAGTCAC-----ACAGTGCTGCAGCCTCAC	1788
Db	264	ThrIleSerTyraSpAsnGluLysGluMetThrGlnLysLeuAspLeuIleThrSerAsp	283
Qy	1789	TCTCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTGACACTTCATCAAAAGATCT	1848
Db	284	MetaLaspTyrglnGlnProLeuMetIleGlyThrValAlaArgLysGlySer	303
Qy	1849	ACCTTTAAACCA-----GAAGA-----GGAAAAGACGACGC	1881
Db	304	ThrPheArgProMetAspThrAspThrGluLysValArgValAsnThrGluAlaSerGly	323
Qy	1882	TATGACAGACCTAGATCTTACAACTCACCAGGCGCAGGAAGTTATCATCCCTATCTGAA	1941
Db	324	His-----TyrAspCysProHisArgProGlyArgHisGluTyraLeu	338
Qy	1942	CCACTCCCAATTACCGGCCTGTAGTATGCACACCCCAATCATC-----	1983
Db	339	ProLeuThrHisSerGluProGluTyraThrProIleValGluArgHisLeuLeuArg	358
Qy	1984	-----ATGCACATGTCAGGCACCCCAACTTCAGTTGGTCAGCCCTCCACA	2031
Db	359	AlaHisThrPheSerThrGlnSerGlyTyraArgValProGly---ProArgProThrHis	377
Qy	2032	TCCACTTCAAGGCTACGGGAACCAACTCCCCACTAGTGGGAAGTTCACATACACTT	2091
Db	378	LysHisSerHisSerSerGlyGlyPheProProAlaThrGlyAlaThrGlnValGluSer	397
Qy	2092	CTCTCCAGGACTACAGCTGCTCTCTCAGCCCGCAGCCAGTATGATACCCCGAAGCTGGG	2151
Db	398	TyrGlnArgProAlaSerProLysProValGlyGlyTyraSplysPro---AlaAla	416
Qy	2152	AGCCAGCTTACCTGCCCCAGCAAGTGGTGTCACCGTGCACAGCAGCACACAAGAA	2211
Db	417	SerSerPheLeuAspSerArgAsp-----ProAlaSerGlnSerGln	430
Qy	2212	GTATCAGGACGAGGAAGGATGGG	2235
Db	431	MetThrSerGlyGlyAspAspGly	438

US-09-864-761-45371
Sequence 45371, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aemica-x-1
CURRENT APPLICATION NUMBER: US/09/864, 761
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27

```
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 45371
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
; OTHER INFORMATION: EST HUMAN HIT: AT565996.1, EVALUE 2.00e-40
; OTHER INFORMATION: SWISSPROT HIT: P98063, EVALUE 1.00e-09
US-09-864-761-45371

Alignment Scores:
Pred. No.: 4e-26 Length: 75
Score: 412.00 Matches: 75
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.94% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-1 (1-2280) x US-09-864-761-45371 (1-75)
QY 145 GATGGATGGACACACTGACTAGGCCCTGAGAGTGGAGACCCCTTACATCCATAAATAC 204
Db 1 AspGlyCysGlyHisThrValLeuGlyProGluSerGlyThrLeuThrSerIleAsnTyr 20
QY 205 CCACAGACCTATCCACACACACTGTTGTGAATGGAGATCCGTGTAAAGATGGGAGAG 264
Db 21 ProGlnThrTyrProAsnSerThrValCysGluTrpGluIleArgValLysMetGlyGlu 40
QY 265 AGAGTTCGCATCAAAATTTGTGACTTGTGACATTGACATTCTGATCTTGTCACTTTAAT 324
Db 41 ArgValArgIleLysPheGlyAspPheAspIleGluAspSerAspSerCysHisPheAsn 60
QY 325 TACTTGAAGATTTATAATGAATTTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 369
Db 61 TyrLeuArgIleTyrAsnGlyIleGlyValSerArgThrGluIle 75

RESULT 6
US-10-104-440-2
; Sequence 2, Application US/1010440

; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Quan
; TITLE OF INVENTION: ANTAGONISTS OF NEUROPILIN RECEPTOR FUNCTION AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 48802 C
; CURRENT APPLICATION NUMBER: US/10/104,440
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/580,803
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/069,155
; PRIOR FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 60/069,687
; PRIOR FILING DATE: 1997-12-29
; PRIOR APPLICATION NUMBER: 60/078,541
; PRIOR FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 2
; LENGTH: 923
; TYPE: PRT
; ORGANISM: human
; US-10-104-440-2

Alignment Scores:
Pred. No.: 7.09e-25 Length: 923
Score: 401.50 Matches: 133
Percent Similarity: 41.41% Conservative: 67
Best Local Similarity: 27.54% Mismatches: 180
Query Match: 9.69% Indels: 103
DB: 12 Gaps: 19

US-10-060-830-1 (1-2280) x US-10-104-440-2 (1-923)
QY 31 CTCTCCGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTC 90
Db 1 MetGluArgGlyLeuPro-----6
QY 91 CTGCTCTTACTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 150
Db 7 LeuLeuCysAlaValLeuAlaLeuValLeuAlaProAlaGlyAlaPheArgAsnAspLys 26
QY 151 TGTGGACACACTGTACTAGGCCCTGAGAGTGGAGACCCCTTACATCCATAAATACCCAG 210
Db 27 CysGlyAspThrIleLysIleGluSerProGlyTyrLeuThrSerProGlyTyrProHis 46
QY 211 ACCTATCCCAACAGCAGCTGTTTGTGAATGGAGATCCCTGTA---AAGATGGAGAGAGA 267
Db 47 SerTyrHisProSerGluLysCysGluTrpLeuIleGlnAlaProAspProTyrGlnArg 66
QY 268 GTTCGCATCAAAATTTGGT---GACTTTGCATTTGAAGATTCTGATTTCTGTCACTTTAAT 324
Db 67 IleMetIleAsnPheAsnProHisPheAspLeuGluAspArgAsp---CysLysTyrAsp 85
QY 325 TACTTGAAGATTTATAATGAATTTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 384
Db 86 TyrValGluValPheAspGlyGluAsnGlyHisPheArgGlyLysPheCys---104
QY 385 CTGGGGTGGCAATGAACCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 444
Db 105 ---GlyLysIleAlaProProValSerSerGlyProPheLeuPheIleLysPhe 123
QY 445 ATGAGTGAATCCCATGTTTCTGGAGCCGGATTTTGGCCCTCATCTCTGTTATAGATAAA 504
Db 124 ValSerAspTyrGluThrHisGlyAlaGlyPheSerIleArgTyrGluIlePheLysArg 143
QY 505 -----CAAGATCTAATAATCT-----519
Db 144 GlyProGluCysSerGluAsnTyrThrThrProSerGlyValIleLysSerProGlyPhe 163
QY 519 -----519
```

```
Db 164 ProGluLysTyrProAsnSerLeuGluCysThrTyrIleValPheAlaProLysMetSer 183
QY 520 -----TGTTTGGACACTGCATCCAAATTTTGGAACTGAGTTCAGTAAGTACTGCCCA 573
Db 184 GluIleLeuGluPheGluSerPheAspLeuGluProAsnSerAsnPro-----Pro 201
QY 574 GCTGGTGTCTGCTCTTTCCTGCTGATATCTCGAACAAATCTCTCATATGATATGATATGAT 633
Db 202 GlyGlyMetPheCysArgTyrAspArgLeuGlu-----IleTyrAspGlyPheProAsp 219
QY 634 TCTCTCCCAATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 681
Db 220 ValGlyPro-----HisIleGlyArgTyrCysGlnLeuGlyThrProGlyArgIleArg 237
QY 682 ACCTGGGGCGGCAATCATGTTGTAATTAAGTAAAGGATATCCCTATTATGATGAAATGCT 741
Db 238 SerSerSerGlyIleLeuSerMetValPhe-----TyrThrAspSerAla 252
QY 742 TTGGCTAAC-----AACGTCACATCTGTGGTGGGACACTTATCTACAAGT 786
Db 253 IleAlaLysGluGlyPheSerAlaAsnTyrSerValLeuGlnSerSerValSerGluAsp 272
QY 787 CTTTTCATATTAAGACAACTGATGTTATGGAACACTGGGATGAGTCTGCTGATC 846
Db 273 -----PheLys-----CysMetGluAlaLeuGlyMetGluSerGlyGluIle 286
QY 847 GCGGATCCTCAATAACAGCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 906
Db 287 HisSerAspGlnIleThrAlaSerSer-----GlnTyrSer 298
QY 907 AACAGTTGGAAACCCAAAAAGCAGGCTGAAAAACCTGGACCGCTTGGGCTGCTGCTTTT 966
Db 299 ThrAsnTyrSerAlaGluArgSerArgLeuAsnTyrProGluAsnGlyTyrThrProGly 318
QY 967 CCACCTGATGAATACAGCTGTTACAAATAGATTTGAAATGAAGAAAGAAATAACAGGC 1026
Db 319 GluAspSerTyrArgGluThrIleGlnValAspLeuGlyLeuLeuArgPheValThrAla 338
QY 1027 ATTATAACCACTGGA-----TCCACCATGCTGGGACACAAATACATGCTGCTGCTGCTAC 1080
Db 339 ValGlyThrGlnGlyAlaIleSerLysLysLysLysLysLysLysLysLysLysLysLys 358
QY 1081 AGAATCCTGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 359 LysIleAspValSerSerAsnGlyGluAspTyrIleThrIleLysGluGly-----Asn 376
QY 1141 CAAGATAAGATATTCAGGAACAAAGATATATCACCAGCATGCTGCTGCTGCTGCTGCTGCTG 1200
Db 377 LysProValLeuPheGlnGlyAsnThrAsnProThrAspValValValAlaValPhePro 396
QY 1201 CCACCAATATTGCACCTTTTATTAGATGAATCCTACCAATGGCAGCAGAAATGGCC 1260
Db 397 LysProLeuIleThrArgPheValArgIleLysProAlaThrIleGluThrGlyIleSer 416
QY 1261 ATGAAATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 417 MetArgPheGluValTyrGlyCys-----LysIleThr 427
QY 1321 CAACCTCCA 1329
Db 428 AspTyrPro 430
RESULT 7
US-10-104-440-4
; Sequence 4, Application US/10104440
; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Quan
; TITLE OF INVENTION: ANTAGONISTS OF NEUROPILIN RECEPTOR FUNCTION AND USE
; TITLE OF INVENTION: THEREOF
```

```
; FILE REFERENCE: 48802 C
; CURRENT APPLICATION NUMBER: US/10/104,440
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/580,803
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/069,155
; PRIOR FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 60/069,687
; PRIOR FILING DATE: 1997-12-29
; PRIOR APPLICATION NUMBER: 60/078,541
; PRIOR FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 931
; TYPE: PRT
; ORGANISM: human
US-10-104-440-4

Alignment Scores:
Pred. No.: 1,25e-24 Length: 931
Score: 398.50 Matches: 199
Percent Similarity: 36.37% Conservative: 112
Best Local Similarity: 23.27% Mismatches: 340
Query Match: 9.62% Indels: 204
DB: 12 Gaps: 36

US-10-060-830-1 (1-2280) x US-10-104-440-4 (1-931)
QY 76 ATCCCTCTGCTCTCTCTG-----CTCTTACTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 129
Db 1 MetAspMetPheProLeuThrTyrPValPheLeuAlaLeuTyrPheSerArgHisGlnVal 20
QY 130 GGAGCCCAAGAGTGTGATGATGTGGACACACTGTACTAGCCCTGAGAGTGGAAACCCCTT 189
Db 21 ArgGlyGlnProAspProProCysGlyGlyArgLeuAsnSerLysAspAlaGlyTyrIle 40
QY 190 ACATCATTAATACACACAGACTATCCCAACAGCAGTGTGTGTGATGGGAGATCCGT 249
Db 41 ThrSerProGlyTyrProGlnAspTyrProSerHisGlnAsnCysGluThrIleValTyr 60
QY 250 GTA---AAGATGGGAGAGAGTTCGCATCAAAATTTGCT---GACTTTTGACATTGAAGAT 303
Db 61 AlaProGluProAsnGlnLysIleValLeuAsnPheAsnProHisPheGluIleGluLys 80
QY 304 TCTGATCTTCTGCTCACTTTAATTAATCTGAGAAATTTATATGGAATTTGAGTGCAGCAACT 363
Db 81 HisAsp---CysLysTyrAspPheIleGluIleArgAspGlyAspSerGluSerAlaAsp 99
QY 364 GAAATAGGCAAAATCTGTGCTGCTGGGTTCGAATGAAC-----CATTCATTAATGAA 414
Db 100 LeuLeuGlyLysHisCysGly-----AsnIleAlaProProThrIleIle 114
QY 415 TCAAAAGCAATGAAATACATCTGCTGCTGATGAGTGAATCCATGTTCTTGGACGCGGA 474
Db 115 SerSerGlySerMetLeuTyrIleLysPheThrSerAspTyrAlaArgGlnGlyAlaGly 134
QY 475 TTTTGGCCCTCAVACTCTGTTTATAGATAAACAAGATCTAATTACTTGTGTGGACACTGCA 534
Db 135 PheSerLeuArgTyrGluIlePheLysThrGlySer-----GluAspCysSer 150
QY 535 TCCAAATTTTGGAACTGATGATTCAGTAAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 594
Db 151 LysAsnPheThrSerProAsnGlyThrIleGluSerProGly-----Phe 165
QY 595 CTGAGATATCTGGAACAATTCCTCATGATGATATAGATATCTCTGCCATTCCTGCTGCTGCT 654
Db 166 ProGlu-----LysTyrProHisAsnLeu---AspCysThrPheThrIleLeuAla 181
QY 655 GGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 702
Db 182 LysProLysMetGluIleIleLeuGlnPheLeuIlePheAspLeuGluHisAspProLeu 201
```

QY	703	-----GTTGCTAATFAGTAAAGTATTCC-----	726
Db	202	GlnValGlyGluGlyAspCysLysTyrAspTrpLeuAspIleTrpAspGlyIleProHis	221
QY	727	-----TATTATGAAGTCTTTGGCTAAACAACGTCACTACTGTG	765
Db	222	ValGlyProLeuIleGlyLysTyrCysGlyThrLysThrProSerGluLeuArgSerSer	241
QY	766	GTGGGACACTTATCTACAAGCTTTTACAA	795
Db	242	ThrGlyIleLeuSerLeuThrPheHisThrAspMetAlaValAlaLysAspGlyPheSer	261
QY	796	-----TTTAAGACAAGTGGATGTTATGGATGAACACATG	825
Db	262	AlaArgTyrTyrLeuValHisGlnGluProLeuGluAsnPheGlnCysAsnValProLeu	281
QY	826	GGATGGAGTCTGGTCTCATCGCGATCCCAATACACACATCACTCTGTCTGGAGTGG	885
Db	282	GlyMetGluSerGlyArgIleAlaAsnGluGlnIleSerAlaSerSer	297
QY	886	ACTGACCACAGGACGAGACAGTGTGGAACCCCAAAACCCAGGCTGAAAAAACCT	945
Db	298	-----ThrTyrSerAspGlyArgTyrThrProGlnGlnSerArgLeuHisGlyAsp	314
QY	946	GGACCGCTTGGTCTGTTTGGCCACTCATGAATACCAGTGGTTACAAATAGATTTGAAT	1005
Db	315	AspAsnGlyTyrThrProAsnLeuAspSerAsnLysGluTyrLeuGlnValAspLeuArg	334
QY	1006	AAGGAAGAAATTAACAGGCATTTATACCACTCGA-----TCCACCATTGGTGGAGCAC	1059
Db	335	PheLeuThrMetIleThrAlaIleAlaThrGlnGlyAlaIleSerArgGluThrGlnAsn	354
QY	1060	AATTACTTGTCTGCTCCCTACAGATCCCTGTACAGTGTATGGCAGAGAATGGACTGTG	1119
Db	355	GlyTyrTyrValLysSerTyrLysLeuGluValSerThrAsnGlyGluAspTrpMetVal	374
QY	1120	TACAGAGACCTGGTGGAGCAAGATAAGATATTTTCAAGAAACAAGATTATCACACAG	1179
Db	375	TyrArgHisGly-----LysAsnHisLysValPheGlnAlaAsnAspAlaThrGlu	392
QY	1180	GATGTCGTAAATACCTTTTGGCCACCAATTTATTCAGCTTTTATTAGACTGATCTACC	1239
Db	393	ValValLeuAsnLysLeuHisAlaProLeuLeuThrArgPheValArgIleArgProGln	412
QY	1240	CAATGCGACAGAAATATGCCATGAATGGAGTGTCTGCGAGTCTCAGTTTATTCCTAAA	1299
Db	413	ThrTrpHisSerGlyIleAlaLeuArgLeuGluLeuPheGlyCys-----	427
QY	1300	GGTCGCTCCAAAACCTTACTCAACCTCCACTCCTCTCGGACACGAATACCTCAAAAAC	1359
Db	428	-----ArgValThrAspAlaProCysSerAsnMetLeuGlyMetLeuSerGly	443
QY	1360	ACTACAGCCCTCCAAAATATACCAAGGTGTGCCCCAAATTT-----	1404
Db	444	LeuIleAlaAspSerGlnIleSerAlaSerSerThrGlnGluTyrLeuTrpSerProSer	463
QY	1405	ACGCAACCACTACACTCGCAGTAGTAATGAATTTCTCTGCACAGACAGACAACAAC	1464
Db	464	AlaAlaArgLeuValSerSerArgSerGlyTrpPheProArgIleProGlnAlaGlnPro	483
QY	1465	GCCAGCTCTGATATCAGAAATACTACCGTAACCTCAATGTAAACCAAGATGTA-----	1518
Db	484	GlyGluGluTrpLeuGlnValAspLeuGlyThrProLysThrValLysGlyValIleIle	503
QY	1519	-----CGCTGGCTCGAGTCTTGTCCCTGTCTGGTGCATCGTCTC	1557
Db	504	GlnGlyAlaArgGlyGlyAspSerIleThrAlaValGluAlaAlaArgAlaPheValArgLys	523
QY	1558	CTCACTACTCTCATTTCTATATTAGTGTGCTGCTGGCACTCG-----AGA	1602
Db	524	PheLysValSerTyrSerLeuAsnGlyLysAspTrpGluTyrIleGlnAspProArgThr	543
QY	1603	AACAGAAAGAAAAAACTCAAGGCACCC-----TATGACTTACCTTACTGCGACCGGCA	1656

Db	544	GlnGlnProLysLeuPheGluGlyAsnMethHisTyrAspThrPro	558
QY	1657	GCTGGTGGAAAGGAATGAAGCAGTTTTCTTCTCTCAAAAGCAGTGGACCATTGAGAAACC	1716
Db	559	-----AspIleargArgPheaspPro	566
QY	1717	COAGTTCGCTATAGCAGCAGCGAAATTAATCACTGAGTCCAAAGAGAAGTCAACCAGATG	1776
Db	567	ProAlaGlnTyrValArgValTyrProGluArgTyrPserProAlaGlyIleGlyMetArg	586
QY	1777	CTGCAGGCTGACTGCGCAGATAT	1830
Db	587	LeuGluValLeuGlyCysAspThrAspSerLysProThrValGluThrLeuGlyPro	606
QY	1831	ACACTTCATCAAGATCTACCTTTAAACCA	1878
Db	607	ThrValLysSerGluGluThrThrProTyrProThrGluGluAlaThrGluCys	626
QY	1879	GGC-----TATGCAGACCTAGATCCTTACAACCTACCGGGCAGGAAG	1921
Db	627	GlyGluAsnCysSerPheGluAspLysAsp-LeuGlnLeuProSerGlyPheAsnCys	646
QY	1922	-----TTTATCATGCCT-----ATGCTGAACCACTCCCAA	1956
Db	646	sAsnPheaspPheLeuGluGluProCysGlyTrpMetTyrAspHisAlaLysTrpLeuAr	666
QY	1957	GGGCGTGAGTATGCAACCCCATCATCAGGACATCTCAGGGCACCCCAAA	2008
Db	666	g-----ThrThrTrpAlaSerSerSerProAsnAspArgTh	679
QY	2008	-----	2008
Db	679	rPheProAspAspArgAsnPheLeuArgLeuGlnSerAspSerGlnArgGluGlyGlnTy	699
QY	2009	-CTTCAGTTGGTCAGCCCTCCACATCCATCTTTCAGGGCTACGGGGAAACCACTCCCCCA	2067
Db	699	rAlaArgLeuIleSerProPro-ValHisLeu-----ProArgSerProv	714
QY	2068	CTAATGGGAACTTACAACTACCTCTCTC-----	2096
Db	714	aiCysMetGluPheGlnTyrGlnAlaThrGlyArgGlyValAlaLeuGlnValVala	734
QY	2097	-----CAGGACTGACAGCTGCTC-----CTCAGCCCGAGCCCAAGTATGATACC	2139
Db	734	rgGluAlaSerGlnGluSerLysLeuLeuTrpValIle-ArgGluAspGlnGlyGlyGlu	753
QY	2140	CGAAAGCTGGGAAGCCAGGTCTACCTGCGCCCAAGCAATTTGGTGTACCAAGGTGCCACAG	2199
Db	754	TrpLysHisGlyArgIleIleLeuProSerTyrAsp---MetGluTyrGlnIle-----	770
QY	2200	AGCACACAGAAGTATCAGGAGCAGGAAGGATGGGGAA	2238
Db	771	ValPheGluGlyValIleGlyLysGlyArgSerGlyGlu	783

RESULTS

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

Db 2216 LeuGluLeuPheGlyCys 2221
RESULT 9
US-10-095-718-4
; Sequence 4, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; FILE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4
Alignment Scores:
Pred. No.: 9,34e-19 Length: 1431
Score: 327.50 Matches: 112
Percent Similarity: 41.36% Conservative: 58
Best Local Similarity: 27.25% Mismatches: 144
Query Match: 7.90% Indels: 97
DB: 12 Gaps: 15
US-10-060-830-1 (1-2280) x US-10-095-718-4 (1-1431)
QY 298 GAAGATTCGATTCCTGTCTACATTAAT----- 324
Db 1050 GluAsnIleHisSerIleHisPheSerGlyHisValPheThrValArgLysLysGluGlu 1069
QY 325 ---TACTTGAAGATTATATAAT----- 342
Db 1070 TyrLysMetAlaValTyrAsnLeuTyrProGlyValPheGluThrValGluMetLeuPro 1089
QY 343 ---GGAAATTCGAGTCAGCAGAACTGAA-----ATAGGCAAAATACCTGTCTGGGGTTG 393
Db 1090 SerGlnValGlyIleTyrArgIleGluCysLeuIleGlyGluHisLeuGlnAlaGlyMet 1109
QY 394 CAATGAACCATCAATTGAATCAAAAGGCAATGAATCAATTCCTGCTCATGAGTGA 453
Db 1110 SerThrLeuPheLeuValTyrSerLysCysGlnThrProLeuGlyMetAlaSerGly 1129
QY 454 -----ATCCATGTTCTCGGACGC----- 471
Db 1130 HisIleArgAspPheGlnIleThrAlaSerGlyGlnIleThrGlyGlnTrpAlaProLysLeu 1149
QY 472 -----GGATTTTGGCCTCATCTCTCTATTAGAT----- 501
Db 1150 AlaArgLeuHisTyrSerGlySerIleAsnAlaTrpSerThrLysAspPropheSerTrp 1169
QY 502 ---AAACAAGATCTAATTACTTGTGTTG-----GACACTGCATCCCAATTTTGGAA 549
Db 1170 IleLysValAspLeuLeuAlaProMetIleIleHisGlyIleMetThrGlnGlyAlaArg 1189
QY 550 CTTGAGTTCTAGTAAGTACTCCAGAGTGGTGTCTCTCTCTCTCTCTCTCTCTCTCTG 609
Db 1190 GlnLysPheSerLeuTyrValSerGlnPheIleIleMetTyrSer---LeuAspGly 1208
QY 610 ACAATTCCTCATGATATAGAGATTCCTCCCAATTTGCTGCTGCTGCTGCTGCTGCTG 669

```

Db 1209 AsnLysTrpHisSerTyrArgGlyAsnSer-----1218
QY 670 GTAGTGTCAACACAGCTGGCGGCCAAATCATGTTGTAATTAGTAAAGGTATT-----723
Db 1219 -----ThrGlyThrLeuMetValPhePheGlyAsnValAspSerGlyIleLysHis1236
QY 724 -----CCCTATTATGAAGTCTTTGGCTGAACACATGCACA-----TCT762
Db 1237 AsnIlePheAsnProProIleLeuAlaGlnTyrIleArgLeuHisProThrHisTyrSer1256
QY 763 GTGGTGGGACACTTATCTCAACAGCTTTTACATTTTAAGACAAAGTGGATTTATGAACA822
Db 1257 IleArgSerThrLeuArgMetGluLeuLeuGlyCysAspPheAsnSerCysSerMetPro1276
QY 823 CTGGGATGAGGCTGTGTGATGCGGATCCTCAATAACACATCATCATGTGCTGGAG882
Db 1277 LeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyrLeu--1295
QY 883 TGGACTGACACACACAGCGGCAAGAGACAGTTGGAAACCCAAAGCCAGCTGAAAAAA942
Db 1296 -----SerSerMetLeuAlaThrTrpSerProSerGlnAlaArgLeuHisLeu1311
QY 943 CCTGGA-----CCGCTTGGGCTGCTTTTGGCCACTGATGATACACAGTGTACAAATA996
Db 1312 GlnGlyArgThrAsnAlaTrpArgProGlnAlaAsnAsnProLysGluTrpLeuGlnVal1331
QY 997 GATTGAATAAGAAAGAAATAACAGGCAATTATACCACTGGATCCACCATGGTGGAG1056
Db 1332 AspPheArgLysThrMetLysValThrGlyIleThrThrGlnGlyValLysSerLeuLeu1351
QY 1057 CACAATTACTATGCTGCTCCTACAGAACCTCTACAGTATGATGGCAGAAATGACT1116
Db 1352 IleSerMetTyrValLysGluPheLeuIleSerSerGlnAspGlyHisAsnTrpThr1371
QY 1117 GTGTACAGAGAGCTGTGTGGAGCAAGATAAGATATTTCAAGCAACAAAGATTATCAC1176
Db 1372 LeuPheLeuGlnAsnGly-----LysValLysValPheGlnGlyAsnArgSerSer1389
QY 1177 CAGATGTGGTAAATACACTTTTGGCCACCAATTTATGCCAGCTTTTATAGAGTAATCCT1236
Db 1390 ThrProValArgAsnArgLeuGluProProLeuValAlaArgTyrValArgLeuHisPro1409
QY 1237 ACCAATGTCAGCAGAAATATGCAATGAGCTGCTGGATGCTCAGTTATTCCT1296
Db 1410 GlnSerTrpAlaHisIleLeuLeuArgLeuGluValLeuGlyCysAsp-----1426
QY 1297 AAAGTGTGCTCTCCAAACTTACTACACCTCCA1329
Db 1427 -----ThrGlnGlnPro1430

RESULT 10
US-10-006-091-1
; Sequence 1, Application US/10006091
; Patent No. US20020102730A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; CURRENT APPLICATION NUMBER: US/10/006,091
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence

US-10-006-091-1
Alignment Scores:
Pred. No.: 3,35e-17 Length: 1438
Score: 308.50 Matches: 66
Percent Similarity: 54.49% Conservative: 31
Best Local Similarity: 37.08% Mismatches: 72
Query Match: 7.45% Indels: 9
DB: 12 Gaps: 3
US-10-060-830-1 (1-2280) x US-10-006-091-1 (1-1438)
QY 760 TCTGTGTGGGACACACTTATCTCAACAGCTTTTACATTTTAAGACAAAGTGGATGTATGGA819
Db 1263 SerIleArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet1282
QY 820 ACATCGGGATGAGTCTGTGTGATGCGGATCCTCAATAACACATCATCATGTGCTG879
Db 1283 ProLeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyrPhe1302
QY 880 GAGTGGACTGACACACAGCGGCAAGAGACAGTTGGAAACCCAAAGCCAGCTGAA939
Db 1303 -----ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis1317
QY 940 AAACCTGGA-----CCGCTTGGGCTGCTTTTGGCCACTGATGATACCAAGTGTACAA993
Db 1318 LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGln1337
QY 994 ATAGATTTGAATAAGAAAGAAATAACAGGCAATTATACCACTGGATCCACCATGGTG1053
Db 1338 ValAspPheGlnLysThrMetLysValThrGlyValThrThrGlnGlyValLysSerLeu1357
QY 1054 GACCAATTAATCTATGCTGCTGCTACAGATCTGTACAGTATGATGGCAGAAATGG1113
Db 1358 LeuThrSerMetTyrValLysGluPheLeuIleSerSerGlnAspGlyHisGlnTrp1377
QY 1114 ACTGTGTACAGAGAGCTGTGTGGAGCAAGATAAGATATTTCAAGCAACAAAGATTAT1173
Db 1378 ThrLeuPhePheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer1395
QY 1174 CACAGATGTGGTAAATACACTTTTGGCCACCAATTTATGCCAGCTTTTATAGAGTAAT1233
Db 1396 PheThrProValValAsnSerLeuAspProLeuLeuThrArgTyrLeuArgIleHis1415
QY 1234 CCTACCAATGGCAGCAGAAATTTGCCATGAAATGGAGCTGCTGGATGTCAG1287
Db 1416 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu1433

RESULT 11
US-10-047-257-1
; Sequence 1, Application US/10047257
; Patent No. US20020115152A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.2
; CURRENT APPLICATION NUMBER: US/10/047,257
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-047-257-1
Alignment Scores:

Pred. No.: 3 35e-17 Length: 1438
Score: 308.50 Matches: 66
Percent Similarity: 54.49% Conservative: 31
Best-Local Similarity: 37.08% Mismatches: 72
Query Match: 7.45% Indels: 9
DB: 12 Gaps: 3

US-10-060-830-1 (1-2280) x US-10-047-257-1 (1-1438)

QY 760 TCTGTGGTGGACACATCTATCTACAACTCTTTTACATTTAAAGACAAGTGGATGTTATGGA 819
|||:|:| ||| ||| :|:|
Db 1263 SerileArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet 1282

QY 820 ACACCTGGGATGCTGTGTGATCGCGGATCTCAAAATACACATCATCTGTGCTG 879
|||:|:| ||| ||| :|:|
Db 1283 ProLeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyPhe 1302

QY 880 GAGTGGACTGACACACAGGCGCAAGAGACAGTGTGAAACCCAAAGCCAGGCTGAAA 939
|||:|:| ||| ||| :|:|
Db 1303 -----ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis 1317

QY 940 AAACCTGGA-----CGGCTTGGCTGCTTTTGGCCACTGATCAATACCAAGTGGTTACAA 993
|||:|:| ||| ||| :|:|
Db 1318 LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGln 1337

QY 994 ATAGATTGTAATAGGAAAGAAATAACAGGCAATATTAACACTGGATCCACCATGGTG 1053
|||:|:| ||| ||| :|:|
Db 1338 ValAspPheGlnLysThrMetLysValThrGlyValThrGlnGlyValLysSerLeu 1357

QY 1054 GAGCACAATTTACTGTCTGCTCCTACAGATCTGTACAGTATGATGGGAGAGAAATGG 1113
|||:|:| ||| ||| :|:|
Db 1358 LeuThrSerMetTyrrValLysGluPheLeuIleSerSerSerGlnAspGlyHisGlnTrp 1377

QY 1114 ACTGTGTACAGACGCTGTGTGTGGAGCAAGATAAGATATTTCAAGGAAACAAGATTAT 1173
|||:|:| ||| ||| :|:|
Db 1378 ThrLeuPheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer 1395

QY 1174 CACCAGATGTGGTATAACTTTTGGCCACCAATTTATGACAGTGTATTAAGAGTGAAT 1233
|||:|:| ||| ||| :|:|
Db 1396 PheThrProValValAsnSerLeuAspProProLeuLeuThrArgTyrrLeuArgIleHis 1415

QY 1234 CCTACCAATGGCAGCAGAAATGCCATAAATGGAGTGTCTCGGATGTCAG 1287
|||:|:| ||| ||| :|:|
Db 1416 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 1433

RESULT 12

US-10-095-718-2

; Sequence 2, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; FILE REFERENCE: Methods of Using the Same
; CURRENT APPLICATION NUMBER: US/10/095,718
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1471
; TYPE: PRT
; ORGANISM: Homo sapiens B-domain deleted factor VIII
; FEATURE:
; OTHER INFORMATION: Homo sapiens BDD FVIII

US-10-095-718-2

Alignment Scores:
Pred. No.: 3 37e-17 Length: 1471
Score: 308.50 Matches: 66
Percent Similarity: 54.49% Conservative: 31
Best-Local Similarity: 37.08% Mismatches: 72
Query Match: 7.45% Indels: 9
DB: 12 Gaps: 3

US-10-060-830-1 (1-2280) x US-10-095-718-2 (1-1471)

QY 760 TCTGTGGTGGACACATCTATCTACAACTCTTTTACATTTAAAGACAAGTGGATGTTATGGA 819
|||:|:| ||| ||| :|:|
Db 1296 SerileArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet 1315

QY 820 ACACCTGGGATGCTGTGTGATCGCGGATCTCAAAATACACATCATCTGTGCTG 879
|||:|:| ||| ||| :|:|
Db 1316 ProLeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyPhe 1335

QY 880 GAGTGGACTGACACACAGGCGCAAGAGACAGTGTGAAACCCAAAGCCAGGCTGAAA 939
|||:|:| ||| ||| :|:|
Db 1336 -----ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis 1350

QY 940 AAACCTGGA-----CGGCTTGGCTGCTTTTGGCCACTGATCAATACCAAGTGGTTACAA 993
|||:|:| ||| ||| :|:|
Db 1351 LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGln 1370

QY 994 ATAGATTGTAATAGGAAAGAAATAACAGGCAATATTAACACTGGATCCACCATGGTG 1053
|||:|:| ||| ||| :|:|
Db 1371 ValAspPheGlnLysThrMetLysValThrGlyValThrGlnGlyValLysSerLeu 1390

QY 1054 GAGCACAATTTACTGTCTGCTCCTACAGATCTGTACAGTATGATGGGAGAGAAATGG 1113
|||:|:| ||| ||| :|:|
Db 1391 LeuThrSerMetTyrrValLysGluPheLeuIleSerSerSerGlnAspGlyHisGlnTrp 1410

QY 1114 ACTGTGTACAGACGCTGTGTGTGGAGCAAGATAAGATATTTCAAGGAAACAAGATTAT 1173
|||:|:| ||| ||| :|:|
Db 1411 ThrLeuPheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer 1428

QY 1174 CACCAGATGTGGTATAACTTTTGGCCACCAATTTATGACAGTGTATTAAGAGTGAAT 1233
|||:|:| ||| ||| :|:|
Db 1429 PheThrProValValAsnSerLeuAspProProLeuLeuThrArgTyrrLeuArgIleHis 1448

QY 1234 CCTACCAATGGCAGCAGAAATTTGCCATAAATGGAGTGTCTCGGATGTCAG 1287
|||:|:| ||| ||| :|:|
Db 1449 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 1466

RESULT 13

US-09-957-641-2

; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
; CURRENT FILING DATE: 2001-09-16
; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-957-641-2

Alignment Scores:

Pred. No.: 3 98e-17 Length: 2332
Score: 308.50 Matches: 66

Percent Similarity: 54.49% Conservative: 31
Best Local Similarity: 37.08% Mismatches: 72
Query Match: 7.45% Indels: 9
DB: Gaps: 3

US-10-060-830-1 (1-2280) x US-09-957-641-2 (1-2332)

QY 760 TCCTGCTGGGACACATTATCTCAAGCTCTTTTACATTTTAAGACAAGTGGATGTTATGGA 819
DB 2157 SerileArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSerMet 2176
QY 820 ACATCGGGATGAGTCTGGTGTGATCGCGGATCCCTCAATAACACATCATCTGTGCTG 879
DB 2177 ProLeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerTyrPhe 2196
QY 880 GAGTGCATGCACACAGCGGCAAGAGACAGTGTGAAACCCCAAAAGCCAGCTGAAA 939
DB 2197 -----ThrAsnMetPheAlaThrTrpSerProSerLysAlaArgLeuHis 2211
QY 940 AAACCTGGA-----CGCGCTTGGCTGCTTTTCCCACTGATCAATACCAAGTGTACAA 993
DB 2212 LeuGlnGlyArgSerAsnAlaTrpArgProGlnValAsnAsnProLysGluTrpLeuGln 2231
QY 994 ATAGATTTGAATAAGGAAAGAAATAACAGGCATTATACCACTGGATCCACCATGGTG 1053
DB 2232 ValAspPheGlnLysThrMetLysValThrGlyValThrThrGlnGlyValLysSerLeu 2251
QY 1054 GAGCACATTTACTATGCTGCTGCTACAGAACTCTGTACAGTATGATGGCGCAGAAATGG 1113
DB 2252 LeuThrSerMetTyrValLysGluPheLeuLeuIleSerSerGlnAspGlyHisGlnTrp 2271
QY 1114 ACTGTGTACAGAGAGCTGCTGGAGCAAGATAAGATATTTCAAGGAAACAAAGATTAT 1173
DB 2272 ThrLeuPheGlnAsnGly-----LysValLysValPheGlnGlyAsnGlnAspSer 2289
QY 1174 CACAGGATGCGGTATATACTTTTGGCCACCAATTATTCACAGTCTTTTATAGAGTAAAT 1233
DB 2290 PheThrProValValAsnSerLeuAspProProLeuLeuThrArgTyrLeuArgIleHis 2309
QY 1234 CCTACCAATGGCAGCAGAAATGGCATGAAATGAGCTGCTCGATGTCAG 1287
DB 2310 ProGlnSerTrpValHisGlnIleAlaLeuArgMetGluValLeuGlyCysGlu 2327

RESULT 14

US-09-992-598-62
Sequence 62, Application US/09992598
Patent No. US20020160384A1

GENERAL INFORMATION:

APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kijavini, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2730PIC20
CURRENT APPLICATION NUMBER: US/09/992,598
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861


```
QY 1339 AAC-----AGCAATGACCTCAAAACACTACAGCCCTCCAAAAATAGCCAAA 1386
||| ||| ||| |||
Db 300 AsnTyrTyrHisArgArgAsnGluMetThrThrThr-AspAspLeuAsp----- 315
||| ||| ||| |||
QY 1387 GGTCTGCCCCAAATTTACGCAACACTACAACTCCAGTAGCAATGAAATTTCTCTGCA 1446
||| ||| ||| |||
Db 316 -----PheLysHisHisAsn-----Ty 321
||| ||| ||| |||

QY 1447 CAGACAGACAACAACCTCCAGTCTGTATCATCAGAAATACTACCG 1492
||||| ||| ||| |||
Db 321 rLysGluMetArgGlnLeuMetLysValValAsnGluMetCysPro 336
||||| ||| ||| |||

RESULT 15
US-09-989-293A-62
; Sequence 62, Application US/09989293A
; Patent No. US20020177164A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC66
; CURRENT APPLICATION NUMBER: US/09/989,293A
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827

; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
```

Alignment Scores:		
Pred. No.:	2.87e-10	756
Score:	222.50	66
Percent Similarity:	46.19%	43
Conservative:		

```

Best Local Similarity: 27.97%      Mismatches: 77
Query Match: 5.37%              Indels: 50
DB: 9                            Gaps: 10

US-10-060-830-1 (1-2280) x US-09-989-293A-62 (1-756)

Qy 823 CTGGGGTGGAGCTGGTGTGATCCGGATCTCAATAACACGATCATCTGCTGCTGGAG 882
|||||:|||||: ||| |||: |||||:|||||
Db 139 LeuGlyLeuGluThrLeuLysIleThrAspPheGlnLeuHisAlaSerThrValLysArg 158
TGG-----ACTGACCACACAGGGCAAGAGACAGTGTGGAAACCCAAAAAGCCAGCGCT 936
:|: ||| |||||: ||| |||: |||||: |||
Db 159 TyrGlyLeuGlyAlaHisArgGlyArgLeuAsnIle-----GlnAlaGlyIle 174
Qy 937 AAAAACCT-----GGACCGCTGGGTGCTTTTGCACCTGATGATACACGAGTGG 987
|||: ||| ||| ||| ||| ||| ||| |||
Db 175 AsnGluAsnAspPheTyrAspGlyAlaThrCysAlaGlyArgAsnAspLeuGlnGlnTrp 194
TTCACAAATAGATTTCAAATAAGCAAAAGAAATAACAGCATATATAACCACTGGATCCACC 1047
:|: |||||: |||: |||: |||||: |||||: |||
Db 195 IleGluValAspAlaArgLeuThrArgPheThrGlyValIleThrGlnGlyArgAsn 214
Qy 1048 ATGTTGGAGCAAAATCTATGTCTGCCTACAGAAATCCTGTACGATGATGATGGCAG 1107
|||: |||: |||: |||: |||: |||: |||: |||
Db 215 SerLeuTrpLeuSerAspTrpValThrSerTyrLysValMetValSerAsnAspSerHis 234
Qy 1108 AATATG---ACTGTCTACAGAGCCTGTGTGGACCAAGATAAGATATTTCAGAGAAAC 1164
||| ||||| ||| ||| ||| ||| ||| ||| |||
Db 235 ThrTrpValThrValLysAsnGlySergly-----AspMetIlePheGluGlyAsn 251
Qy 1165 AAGATATTACACAGGATGTGCGTAAATACTTTTGGCCACCAATATTGACACGTTTATT 1224
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 SerGluLysGluIleProValLeuAsnGluLeuProValProMetValAlaArgTyrIle 271
Qy 1225 AGAGTGAATCTTACCCAAATGG-----CAGCAGAAATTCGCCAATGAATAATGGAGCTGCTC 1278
|||||:||||| ||| ||| |||||:|||||: |||
Db 272 ArgIleAsnProGlnSerTrpPheAspAsnGlySerIleCysMetArgMetGluIleLeu 291
Qy 1279 GGATCTCAGTTTATTCTTAAAGTCGTCTCTCCAAAACCTTACTCACTCCACCTCCTCGG 1338
||||| ||||| ||||| ||||| ||||| |||||
Db 292 GlyCysProLeu-----ProAspProAsn 299
Qy 1339 AAC-----AGCAATGACCTCAAAAACACTACAGCCCTCCAAAATAGCCAAA 1386
||| ||| |||: ||| ||| ||| ||| |||
Db 300 AsnTyrTyrHisArgArgAsnGluMetThrThrThr-AspAspLeuAsp----- 315
Qy 1387 GGTCTGCCCCAAATTTACGCANCCACACTACACCTCGCAGTAGCAATGAATTCCTGCA 1446
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 -----PheLysHisAsn-----Tyr 321
Qy 1447 CAGACAGAAACAAACCTGCCAGTCTCGATATCAGAAATACTACCG 1492
|||: |||: |||: |||: |||: |||: |||
Db 321 rlysgluMetArgGlnLeuMetLysValValAsnGluMetCysPro 336

Search completed: January 21, 2003, 09:51:52
Job time : 31.2799 secs

```

Search completed: January 21, 2003, 09:51:52Z
Job time : 31.2799 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:55:46 ; Search time 27.4298 Seconds
(without alignments)
5476.098 Million cell updates/sec

Title: US-10-060-830-3

Perfect score: 729

Sequence: 1 MPLFLLLLVLLLELDAGA.....TOEVSAGAGRGDCDVFKEIL 729

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	729	100.0	775	4 Q96PD2	Q96pd2 homo sapien
2	729	100.0	775	4 Q8TDX2	Q8tdx2 homo sapien
3	84	11.5	364	4 Q14089	Q14089 homo sapien
4	51	7.0	251	11 Q9D9K5	Q9d9k5 mus musculus
5	51	7.0	769	11 Q912V3	Q912v3 mus musculus
6	44	6.0	769	11 Q912V2	Q912v2 rattus norv
7	16	2.2	34	4 Q96IX0	Q96ix0 homo sapien
8	9	1.2	73	2 Q9EUA9	Q9eua9 helicobacte
9	9	1.2	73	2 Q9F6T0	Q9ft60 helicobacte
10	9	1.2	73	2 Q9F6S9	Q9f6s9 helicobacte
11	9	1.2	73	2 Q9F6S8	Q9f6s8 helicobacte
12	9	1.2	155	4 Q96P81	Q96p81 homo sapien
13	9	1.2	237	8 Q8WEB2	Q8web2 mabuya dela
14	9	1.2	237	8 Q8W8F8	Q8w8f8 mabuya dela
15	9	1.2	276	16 Q25256	Q25256 helicobacte
16	9	1.2	296	2 Q51756	Q51756 pseudomonas

17	9	1.2	332	8 Q9B179	Q9b179 mabuya dela
18	9	1.2	332	8 Q9B201	Q9b201 mabuya dela
19	9	1.2	332	8 Q9B200	Q9b200 mabuya dela
20	9	1.2	377	4 Q8TAG4	Q8tag4 homo sapien
21	9	1.2	728	16 Q9HY06	Q9hyq6 pseudomonas
22	9	1.2	1197	13 Q90478	Q90478 brachydanio
23	8	1.1	37	4 Q8TD81	Q8td81 homo sapien
24	8	1.1	71	10 Q8W483	Q8w483 arabidopsis
25	8	1.1	95	10 Q9SKJ8	Q9skj8 arabidopsis
26	8	1.1	95	10 Q39066	Q39066 arabidopsis
27	8	1.1	103	11 Q9CUI3	Q9cui3 mus musculus
28	8	1.1	104	17 Q9UWS7	Q9uws7 sulfolobus
29	8	1.1	112	10 Q9FIC4	Q9fig4 arabidopsis
30	8	1.1	125	5 Q9UIP6	Q9uip6 caenorhabdi
31	8	1.1	139	16 Q92JV2	Q92jv2 rhizobium m
32	8	1.1	157	5 Q8T6R9	Q8t6r9 anopheles g
33	8	1.1	161	5 Q16655	Q16655 caenorhabdi
34	8	1.1	163	16 Q92PM3	Q92pm3 rhizobium m
35	8	1.1	201	17 Q58922	Q58922 pyrococcus
36	8	1.1	203	8 Q9MIB5	Q9mib5 mabuya eleg
37	8	1.1	222	16 Q8YN39	Q8yn39 anabaena sp
38	8	1.1	236	2 Q87318	Q87318 mycobacteri
39	8	1.1	245	11 Q9DIX2	Q9dix2 mus musculus
40	8	1.1	254	4 Q8WXV2	Q8wxv2 homo sapien
41	8	1.1	257	10 Q94CK4	Q94ck4 arabidopsis
42	8	1.1	258	16 Q9Z8D8	Q9z8d8 chlamydia p
43	8	1.1	271	10 Q8VZP2	Q8vzp2 arabidopsis
44	8	1.1	277	10 Q9FMJ5	Q9fmj5 arabidopsis
45	8	1.1	280	5 Q9V166	Q9v166 drosophila

ALIGNMENTS

RESULT 1

Q96PD2	ID	Q96PD2	PRELIMINARY;	PRT;	775 AA.
AC	Q96PD2;				
DT	01-DEC-2001	(TREMBlrel. 19, Created)			
DT	01-DEC-2001	(TREMBlrel. 19, Last sequence update)			
DT	01-JUN-2002	(TREMBlrel. 21, Last annotation update)			
DE	Endothelial and smooth muscle cell-derived neuropilin-like protein.				
DE	protein.				
GN	ESDN.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	PubMed=1144734;				
RA	Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,				
RA	Matsumori A., Sasayama S., Honjo T., Tashiro K.;				
RT	"ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular				
RT	Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is				
RT	Up-regulated after Vascular Injury."				
CC	J. Biol. Chem. 276:34105-34114(2001).				
CC	-!- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL				
CC	REGULATORS.				
DR	EMBL: AF387547; AAL30178.1; -.				
DR	InterPro: IPR000859; CUB_domain.				
DR	InterPro: IPR000421; FA58_C.				
DR	InterPro: IPR004043; LCCL_dom.				
DR	Q912v2 mus musculus				
DR	Pfam: PF00431; CUB; 1.				
DR	Pfam: PF00754; F5_F8_type_C; 1.				
DR	PROSITE: PS01180; CUB; 1.				
KW	DNA-binding; Transcription regulation.				
SO	SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;				

Query Match

Best Local Similarity 100.0%; Score 729; DB 4; Length 775;

Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MPEFLLLVLLVLLLEDAGAGCGGCGHTVLGPESGTLTSINYPTYNSTVCWEIRVK 60
DB 47 MPEFLLLVLLVLLLEDAGAGCGGCGHTVLGPESGTLTSINYPTYNSTVCWEIRVK 106
QY 61 MGERVRIKFGDFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGGLQNMNHSIESKGNIT 120
DB 107 MGERVRIKFGDFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGGLQNMNHSIESKGNIT 166
QY 121 LLFMSGIHVSGRGFLASYVIDKQDLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 180
DB 167 LLFMSGIHVSGRGFLASYVIDKQDLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 226
QY 181 PHGYRDSPLCMAGVHAGVSNLTGGQISVVISKGIPIYESLANNVTSVGHLSLFT 240
DB 227 PHGYRDSPLCMAGVHAGVSNLTGGQISVVISKGIPIYESLANNVTSVGHLSLFT 286
QY 241 FKTSCYCTLGMSGVIAADPOITASSVLEWTDHTGOENSWPKPKARLKKPGPPWAAAFATD 300
DB 287 FKTSCYCTLGMSGVIAADPOITASSVLEWTDHTGOENSWPKPKARLKKPGPPWAAAFATD 346
QY 301 EYQWLQIDLNEKKITGIIITGSTWVEHNYVYSAYRIYSDDGQKWTVYRFGVEQDKIF 360
DB 347 EYQWLQIDLNEKKITGIIITGSTWVEHNYVYSAYRIYSDDGQKWTVYRFGVEQDKIF 406
QY 361 QGNKDYHODVRNFPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPKLTQPPPP 420
DB 407 QGNKDYHODVRNFPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPKLTQPPPP 466
QY 421 RNSDLKNTTAPPKTAKGRAPKFTQLOPRSSNEPQTEQTTASPDIRNTVTNVTKD 480
DB 467 RNSDLKNTTAPPKTAKGRAPKFTQLOPRSSNEPQTEQTTASPDIRNTVTNVTKD 526
QY 481 VALAAVLPVLMVLTLLILVCAWHNRNKKTEGTYDLPYWDRAWMKMKQFLPAK 540
DB 527 VALAAVLPVLMVLTLLILVCAWHNRNKKTEGTYDLPYWDRAWMKMKQFLPAK 586
QY 541 AVDHETPVRYSSSEVNHLSPREVTTLQADSAEYAQPLVGGIVGTLHORSTFKPEEGKE 600
DB 587 AVDHETPVRYSSSEVNHLSPREVTTLQADSAEYAQPLVGGIVGTLHORSTFKPEEGKE 646
QY 601 AGYADLPYNSPGQEVYHAYAEPLITGPEYATPIIMDSHGHTTSVGQPSSTFKATGN 660
DB 647 AGYADLPYNSPGQEVYHAYAEPLITGPEYATPIIMDSHGHTTSVGQPSSTFKATGN 706
QY 661 QPPLVGTNTLLSRDSCSSAAQYDTPKAGKPGLPAPDELVIYVQVQSTQEVSGAGRDG 720
DB 707 QPPLVGTNTLLSRDSCSSAAQYDTPKAGKPGLPAPDELVIYVQVQSTQEVSGAGRDG 766
QY 721 ECDVFEKIL 729
DB 767 ECDVFEKIL 775

RESULT 2
Q8TDX2 ID Q8TDX2 PRELIMINARY; PRT: 775 AA.
AC Q8TDX2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE CLCPl.
DE CLCPl.
GN CLCPl.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC TISSUE=LUNG;
RA Koshikawa K., Osada H., Kozaki K., Konishi H., Masuda A.,
RA Tatematsu Y., Mitsudomi T., Nakao A., Takahashi T.;
RT "Significant up-regulation of a novel gene, CLCPl, in a highly
RT metastatic lung cancer subline as well as in lung cancers in vivo."

```

```

RL Oncogene 0:0-0(2002).
DR EMBL; AB073146; BAB91138.1; - 6D07223B21BE5A42 CRC64;
SQ SEQUENCE 775 AA; 85073 MW; 6D07223B21BE5A42 CRC64;

Query Match 100.0%; Score 729; DB 4; Length 775;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEFLLLVLLVLLLEDAGAGCGGCGHTVLGPESGTLTSINYPTYNSTVCWEIRVK 60
DB 47 MPEFLLLVLLVLLLEDAGAGCGGCGHTVLGPESGTLTSINYPTYNSTVCWEIRVK 106
QY 61 MGERVRIKFGDFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGGLQNMNHSIESKGNIT 120
DB 107 MGERVRIKFGDFDIEDSDSCHFNLYRIYNGIGVSRTEIGKYCGGLQNMNHSIESKGNIT 166
QY 121 LLFMSGIHVSGRGFLASYVIDKQDLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 180
DB 167 LLFMSGIHVSGRGFLASYVIDKQDLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 226
QY 181 PHGYRDSPLCMAGVHAGVSNLTGGQISVVISKGIPIYESLANNVTSVGHLSLFT 240
DB 227 PHGYRDSPLCMAGVHAGVSNLTGGQISVVISKGIPIYESLANNVTSVGHLSLFT 286
QY 241 FKTSCYCTLGMSGVIAADPOITASSVLEWTDHTGOENSWPKPKARLKKPGPPWAAAFATD 300
DB 287 FKTSCYCTLGMSGVIAADPOITASSVLEWTDHTGOENSWPKPKARLKKPGPPWAAAFATD 346
QY 301 EYQWLQIDLNEKKITGIIITGSTWVEHNYVYSAYRIYSDDGQKWTVYRFGVEQDKIF 360
DB 347 EYQWLQIDLNEKKITGIIITGSTWVEHNYVYSAYRIYSDDGQKWTVYRFGVEQDKIF 406
QY 361 QGNKDYHODVRNFPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPKLTQPPPP 420
DB 407 QGNKDYHODVRNFPPIIARFIRVNPQWQOKIAMKMLLGCQFIPKGRPKLTQPPPP 466
QY 421 RNSDLKNTTAPPKTAKGRAPKFTQLOPRSSNEPQTEQTTASPDIRNTVTNVTKD 480
DB 467 RNSDLKNTTAPPKTAKGRAPKFTQLOPRSSNEPQTEQTTASPDIRNTVTNVTKD 526
QY 481 VALAAVLPVLMVLTLLILVCAWHNRNKKTEGTYDLPYWDRAWMKMKQFLPAK 540
DB 527 VALAAVLPVLMVLTLLILVCAWHNRNKKTEGTYDLPYWDRAWMKMKQFLPAK 586
QY 541 AVDHETPVRYSSSEVNHLSPREVTTLQADSAEYAQPLVGGIVGTLHORSTFKPEEGKE 600
DB 587 AVDHETPVRYSSSEVNHLSPREVTTLQADSAEYAQPLVGGIVGTLHORSTFKPEEGKE 646
QY 601 AGYADLPYNSPGQEVYHAYAEPLITGPEYATPIIMDSHGHTTSVGQPSSTFKATGN 660
DB 647 AGYADLPYNSPGQEVYHAYAEPLITGPEYATPIIMDSHGHTTSVGQPSSTFKATGN 706
QY 661 QPPLVGTNTLLSRDSCSSAAQYDTPKAGKPGLPAPDELVIYVQVQSTQEVSGAGRDG 720
DB 707 QPPLVGTNTLLSRDSCSSAAQYDTPKAGKPGLPAPDELVIYVQVQSTQEVSGAGRDG 766
QY 721 ECDVFEKIL 729
DB 767 ECDVFEKIL 775

RESULT 3
Q14089 ID Q14089 PRELIMINARY; PRT: 364 AA.
AC Q14089;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 40.0 kDa protein (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

Query Match	7.08; Score 51; DB 11; Length 251;
-------------	------------------------------------

RA	Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT	"ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT	Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT	Up-regulated after Vascular Injury. ";
RL	J. Biol. Chem. 276:34105-34114(2001).
CC	-1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC	REGULATORS.
DR	ENBL; AF387549; AAL30180.1; -.
DR	InterPro: IPR000859; CUB_domain.
DR	InterPro: IPR000421; FA58_C.
DR	InterPro: IPR004043; LCCL_dom.
DR	Pfam: PF00431; CUB; 1.
DR	Pfam: PF00754; F5_F8_type_C; 1.
DR	PROSITE; PS01180; CUB; 1.
KW	DNA-binding; Transcription regulation.
SW	SEQUENCE 769 AA; 83915 MW; 5B7936C8FA063A93 CRC64;

```
Query Match      6.0%; Score 44; DB 11; Length 769;
Best Local Similarity 100.0%; Pred. No. 4.7e-38;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 558 HLSPREVTTVLQADSAEYAQPLVGGIVGTILHQRTFKPEEGKEA 601
|||||
Db 599 HLSPREVTTVLQADSAEYAQPLVGGIVGTILHQRTFKPEEGKEA 642

RESULT 7		
Q961X0	PRELIMINARY;	PRT; 34 AA.
ID	Q961X0	
AC	Q961X0;	
DT	01-DEC-2001 (TReMBLrel. 19, Created)	
DT	01-DEC-2001 (TReMBLrel. 19, Last sequence update)	
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)	
DE	Unknown (Protein for IMAGE:433276) (Fragment).	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI TaxID=9606;	

11]	SEQUENCE FROM N.A.
RP	TISSUE=SKIN;
RC	Strausberg R.;
RA	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RL	EMBL; BC007117; AAH07117.1; -.
DR	NON_TER 1
FT	SEQUENCE 34 AA; 3533 MW; E6EA7A2239151709 CRC64;
SQ	

```
Query Match      2.2%; Score 16; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 5.1e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	7	LLLVLLEDDAGAAQ	22
		IIIIIIIIIIIIIIII	
Db	1	LLLVLLEDDAGAAQ	16

```

RESULT 8.
Q9EUA9
ID Q9EUA9 PRELIMINARY; PRT; 73 AA.
AC Q9EUA9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HP0519-like protein [Fragment].
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OC NCBI_TaxID=210;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CR41, CR45, AND CR10;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;

```

RT "Composition and gene expression of the cag pathogenicity island in
RT *Helicobacter pylori* strains isolated from gastric carcinoma and
RT gastritis patients in Costa Rica.";
RL submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR	EMBL; AF289396;	AAG09847.1;
DR	EMBL; AF289392;	AAG09848.1;
DR	EMBL; AF289393;	AAG09837.1;
DR	EMBL; AF289397;	AAG09845.1;
FT	NON_TER	73
SQ	SEQUENCE	73 AA: 8853 MW: 1

Query Match	1.2%	Score 9;	DB 2;	Length 73;
Best Local Similarity	100.0%;	Pred. No. 0.38;		
Matches	9;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
			0;	

Qy	5	LLLLVLLL	13
Db	16	LLLLVLLL	24

RESULT	9
Q9F6T0	
ID	Q9F6T0
AC	PRELIMINARY;
DT	PRT; 73 AA.
DT	01-MAR-2001 (T-EMBLrel. 16, Created)
DT	01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT	01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE	HP0519-like protein (fragment).
OS	Helicobacter pylori (Campylobacter pylori).
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC	Helicobacter.
OX	NCBI_TaxID=210;

RP	SEQUENCE FROM N.A.	
RC	STRAIN=CR46;	
RX	MEDLINE=21117015; PubMed=11179371;	
RA	Occhialini A., Marais A., Urdaci M., Sierra R., Covacci A.,	
RA	Meagall F.;	
RT	"Composition and Gene Expression of the cag Pathogenicity Island in	
RT	Helicobacter pylori Strains Isolated from Gastric Carcinoma and	
RT	Gastritis Patients in Costa Rica.";	
RL	Infect. Immun. 69:1902-1908(2001).	
DR	EMBL; AF289394; AAG09839.1; -.	
DR	NON_TER	73
DR		73
DR	SEQUENCE	73 AA; 8819 MW; 1BD644AF5583BF654 CRC64;
DR		

Query Match	1.2%	Score 9;	DB 2;	Length 73;
Best Local Similarity	100.0%;	Pred. No. 0.38;		
Matches 9;	Conservative	0;	Mismatches	0;
Indels				0;
Gaps				0;

Qy	5	LLLLVLLL	13
Db	16	LLLLVLLL	24

RESULT 10	Q9F6S9	PRELIMINARY;	PRT;	73 AA.
ID	Q9F6S9			
AC	Q9F6S9			
DT	01-MAR-2001	(TREMBLrel. 16, Created)		
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	HP0519-like protein (Fragment).			
DE	Helicobacter pylori (Campylobacter pylori).			
OS	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;			
OC	Helicobacter.			
OX	NCBI_Taxid=210;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=CR47.			
RX	MEDLINE=21117015; PubMed=11179371;			
RA	Occichialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci			
RA	Mcgraud F.;			

*Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.*;
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL: AF289395; AAG09841.1; -;
FT NON_TER 73 73
SQ SEQUENCE 73 AA; 8818 MW; 07B4D3B81224EASE CRC64;

Query Match 1.2%; Score 9; DB 2; Length 73;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLL 13
Db 16 LLLLLVLL 24

RESULT 11
Q9F6S8 PRELIMINARY; PRT; 73 AA.
AC Q9F6S8;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HP0519-like protein (Fragment).
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR39;
RX MEDLINE=21117015; PubMed=11179371;
RA Occhialini A., Marais A., Urdaci M., Sierra R., Munoz N., Covacci A.,
RA Megraud F.;
RT "Composition and Gene Expression of the cag Pathogenicity Island in
RT Helicobacter pylori Strains Isolated from Gastric Carcinoma and
RT Gastritis Patients in Costa Rica.*;
RL Infect. Immun. 69:1902-1908(2001).
DR EMBL: AF289396; AAG09843.1; -;
FT NON_TER 73 73
SQ SEQUENCE 73 AA; 8817 MW; 07B4D918B824EASE CRC64;

Query Match 1.2%; Score 9; DB 2; Length 73;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLL 13
Db 16 LLLLLVLL 24

RESULT 12
Q9G6P81 PRELIMINARY; PRT; 155 AA.
AC Q9G6P81;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PTPRE (Fragment).
GN PTPRE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Zhao Y., Sun D., Dai S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF406357; AAL01375.1; -;
FT NON_TER 155 155
SQ SEQUENCE 155 AA; 17359 MW; 8A7DB29677F06E00 CRC64;

Query Match 1.2%; Score 9; DB 4; Length 155;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLL 13
Db 56 LLLLLVLL 64

RESULT 13
Q8W8F8 PRELIMINARY; PRT; 237 AA.
AC Q8W8F8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Cytochrome b (Fragment).
GN CyTB.
OS Mabuya delalandii.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Scincoidae;
OC Scincidae; Mabuya.
OX NCBI_TaxID=149020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=67WDEST;
RA Carranza S., Arnold N., Mateo J., Lopez-Jurado L.;
RT "Parallel gigantism and complex colonization patterns in Cape Verde
RT scincid lizards Mabuya and Macroscincus (Reptilia: Scincidae) revealed
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF280320; AAL55225.1; -;
DR InterPro: IPR000179; CyT.b.b6.
DR Pfam: PF00033; cytochrome_b_N; 1.
DR PROSITE: P500192; CYTOCHROME_B_HEME; UNKNOWN_1.
KW Mitochondrion.
FT NON_TER 1 1
FT NON_TER 237 237
SQ SEQUENCE 237 AA; 26439 MW; 73A5BADDBA7F7D9A CRC64;

Query Match 1.2%; Score 9; DB 8; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LLLLLVLL 14
Db 202 LLLLLVLL 210

RESULT 14
Q8W8F8 PRELIMINARY; PRT; 237 AA.
AC Q8W8F8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Cytochrome b (Fragment).
GN CyTB.
OS Mabuya delalandii.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Scincomorpha; Scincoidae;
OC Scincidae; Mabuya.
OX NCBI_TaxID=149020;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VARIOUS STRAINS;
RA Carranza S., Arnold N., Mateo J., Lopez-Jurado L.;
RT "Parallel gigantism and complex colonization patterns in Cape Verde
RT scincid lizards Mabuya and Macroscincus (Reptilia: Scincidae) revealed
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF280320; AAL55225.1; -;
DR InterPro: IPR000179; CyT.b.b6.
DR Pfam: PF00033; cytochrome_b_N; 1.
DR PROSITE: P500192; CYTOCHROME_B_HEME; UNKNOWN_1.
KW Mitochondrion.
FT NON_TER 1 1
FT NON_TER 237 237
SQ SEQUENCE 237 AA; 26439 MW; 73A5BADDBA7F7D9A CRC64;

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF280315; AAL55220.1; -
 DR EMBL; AF280316; AAL55221.1; -
 DR EMBL; AF280317; AAL55222.1; -
 DR EMBL; AF280319; AAL55224.1; -
 DR EMBL; AF280321; AAL55226.1; -
 DR EMBL; AF280322; AAL55227.1; -
 DR EMBL; AF280323; AAL55228.1; -
 DR EMBL; AF280324; AAL55229.1; -
 DR EMBL; AF280325; AAL55230.1; -
 DR EMBL; AF280326; AAL55231.1; -
 DR EMBL; AF280327; AAL55232.1; -
 DR InterPro; IPR00179; Cyt_b_b6.
 DR Pfam; PF00033; cytochrome_b_N; 1.
 DR PROSITE; PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
 KW Mitochondrion.
 FT NON_TER 1 237
 FT SEQUENCE 237 AA; 26455 MW; 7297BADDBA7F7D84 CRC64;

Query Match 1.2%; Score 9; DB 8; Length 237;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 LLLLVLLL 14
 Db 202 LLLLVLLL 210

RESULT 15

O25256
 ID O25256 PRELIMINARY; PRT; 276 AA.
 AC O25256;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein HP0519.
 GN HP0519.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 pylori.";
 RL Nature 388:539-547(1997).
 RL EMBL; AE000566; AAD07587.1; -
 DR TIGR; HP0519;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 276 AA; 31527 MW; 7B987D1860EF9743 CRC64;

Query Match 1.2%; Score 9; DB 16; Length 276;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLVLLL 13
 Db 16 LLLLVLLL 24

Search completed: January 21, 2003, 09:59:29

Job time : 31.4298 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:55:15 ; Search time 12.1324 seconds
(without alignments)
1767.934 Million cell updates/sec

Title: US-10-060-830-3

Perfect score: 729

Sequence: 1 MPLEFLLLLVLLLLLLEADAGA.....TQEVSGAGRGDCDVFREIL 729

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pap.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pap.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pap.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pap.*
- 5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pap.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	1.1	113	4	US-08-816-772-9
2	8	1.1	114	1	US-08-352-324A-3
3	8	1.1	114	2	US-08-862-607-3
4	8	1.1	114	2	US-08-468-819-4
5	8	1.1	114	3	US-09-203-235-3
6	8	1.1	114	4	US-08-679-493A-149
7	8	1.1	114	5	PCT-US95-16144-3
8	8	1.1	117	4	US-09-134-001C-5606
9	8	1.1	349	2	US-09-032-523-3
10	8	1.1	358	2	US-08-748-485-6
11	8	1.1	396	4	US-08-208-007A-13
12	8	1.1	396	4	US-09-032-523-9
13	8	1.1	396	4	US-08-915-095A-13
14	8	1.1	396	4	US-08-798-096-13
15	8	1.1	396	4	US-08-798-095A-13
16	8	1.1	480	2	US-08-828-488-8
17	8	1.1	480	4	US-09-299-689A-8
18	8	1.1	574	2	US-08-756-317-4
19	8	1.1	600	2	US-08-756-317-3
20	8	1.1	638	2	US-08-756-317-2
21	8	1.1	1503	4	US-08-976-255-14
22	7	1.0	15	4	US-09-336-536-5
23	7	1.0	19	3	US-08-448-194-12
24	7	1.0	19	4	US-08-867-921-12
25	7	1.0	30	4	US-08-448-489-7
26	7	1.0	100	3	US-08-476-376-2
27	7	1.0	108	4	US-08-654-482-4

28 7 1.0 109 4 US-09-071-035-194 Sequence 194, Appl
29 7 1.0 135 4 US-09-288-143-69 Sequence 69, Appl
30 7 1.0 144 2 US-08-341-843B-27 Sequence 27, Appl
31 7 1.0 144 2 US-08-421-497E-32 Sequence 32, Appl
32 7 1.0 150 2 US-08-867-676-3 Sequence 3, Appl
33 7 1.0 151 2 US-08-620-694A-8 Sequence 8, Appl
34 7 1.0 151 3 US-09-034-810-6 Sequence 6, Appl
35 7 1.0 151 3 US-09-022-255-8 Sequence 8, Appl
36 7 1.0 151 3 US-09-022-255-8 Sequence 8, Appl
37 7 1.0 151 3 US-08-685-239-6 Sequence 6, Appl
38 7 1.0 151 3 US-09-022-253-8 Sequence 8, Appl
39 7 1.0 151 3 US-09-022-260-8 Sequence 8, Appl
40 7 1.0 151 4 US-09-022-259-8 Sequence 8, Appl
41 7 1.0 151 4 US-09-022-257-8 Sequence 8, Appl
42 7 1.0 151 4 US-08-432-994A-4 Sequence 4, Appl
43 7 1.0 161 2 US-08-286-819A-25 Sequence 25, Appl
44 7 1.0 161 3 US-08-980-357-25 Sequence 25, Appl
45 7 1.0 163 2 US-08-867-676-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-816-772-9
; Sequence 9, Application US/08816772
; Patent No. 6410268
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
; APPLICANT: LI, HAODONG
; APPLICANT: SU, JEFFREY
; TITLE OF INVENTION: CHEMOKINE ALPHA-3
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,772
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0980001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-816-772-9

Query Match 1.1%; Score 8; DB 4; Length 113;

Best Local Similarity 100.0%; Pred. No. 7.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15

Db 21 LLVLLLLL 28

```

RESULT 2
US-08-352-324A-3
; Sequence 3, Application US/08352324A
; Patent No. 5633149
; GENERAL INFORMATION:
; APPLICANT: Guegler, Karl J.
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Sellhauer, Jeffrey J.
; TITLE OF INVENTION: A NOVEL CHEMOKINE EXPRESSED IN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/352,324A
; FILING DATE: 07-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0025 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 114 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-352-324A-3

Query Match 1.1%; Score 8; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
Db 21 LLVLLLLL 28

RESULT 3
US-08-862-607-3
; Sequence 3, Application US/08862607
; Patent No. 5844084
; GENERAL INFORMATION:
; APPLICANT: Guegler, Karl J.
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Sellhauer, Jeffrey J.
; TITLE OF INVENTION: A NOVEL CHEMOKINE EXPRESSED IN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,607
; FILING DATE: 23-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/352,324
; FILING DATE: 07-DEC-1994
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0025 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 114 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-862-607-3

Query Match 1.1%; Score 8; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
Db 21 LLVLLLLL 28

RESULT 4
US-08-468-819-4
; Sequence 4, Application US/08468819
; Patent No. 5871723
; GENERAL INFORMATION:
; APPLICANT: Strieter, Robert M.
; APPLICANT: Polverini, Peter J.
; APPLICANT: Kunkel, Steven L.
; TITLE OF INVENTION: CXK Chemokines as Regulators of
; TITLE OF INVENTION: Angiogenesis
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,819
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UMIC:003/HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7477
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 4:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 114 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-468-819-4

Query Match 1.1%; Score 8; DB 2; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LVLVLLLL 15
Db 21 LVLVLLLL 28

RESULT 5

US-09-203-235-3
; Sequence 3, Application US/09203235
; Patent No. 6071701
; GENERAL INFORMATION:
; APPLICANT: Guegler, Karl J.
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Wilde, Craig G.
; APPLICANT: Seilhammer, Jeffrey J.
; TITLE OF INVENTION: A NOVEL CHEMOKINE EXPRESSED IN
; TITLE OF INVENTION: INFLAMED ADENOID, ITS PRODUCTION AND USES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/203,235
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/862,607
; FILING DATE: 23-MAY-1997
; APPLICATION NUMBER: 08/352,324
; FILING DATE: 07-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0025 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 114 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-203-235-3

Query Match 1.1%; Score 8; DB 3; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LVLVLLLL 15
Db 21 LVLVLLLL 28

RESULT 6

US-08-679-493A-149
; Sequence 149, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: US/08/679,493A
; CURRENT FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 149
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-679-493A-149

Query Match 1.1%; Score 8; DB 4; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LVLVLLLL 15
Db 21 LVLVLLLL 28

RESULT 7

PCT-US95-16144-3
; Sequence 3, Application PC/TUS9516144
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; TITLE OF INVENTION: A NOVEL CHEMOKINE EXPRESSED IN INFLAMED
; TITLE OF INVENTION: ADENOID, ITS PRODUCTION AND USES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16144
; FILING DATE: 07-DEC-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/352,324
; FILING DATE: 07-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: LUTHER, BARBARA J.
; REGISTRATION NUMBER: 33954
; REFERENCE/DOCKET NUMBER: PF-0025 PCT
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 114 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```
; MOLECULE TYPE: protein
PCT-US95-16144-3

Query Match      1.1%; Score 8; DB 5; Length 114;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
Db 21 LLVLLLLL 28

RESULT 8
US-09-134-001C-5606
; Sequence 5606, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5606
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5606

Query Match      1.1%; Score 8; DB 4; Length 117;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LFLVLLLV 10
Db 32 LFLVLLLV 39

RESULT 9
US-09-032-523-3
; Sequence 3, Application US/09032523
; Patent No. 6232454
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl
; APPLICANT: Baugh, Mariah
; TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,523
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

Query Match      1.1%; Score 8; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLL 12
Db 5 LLLLLVLL 12

RESULT 10
US-08-748-485-6
; Sequence 6, Application US/08748485
; Patent No. 5817480
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Guegler, Karl J.
; APPLICANT: Goli, Surya K.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: NOVEL HISTAMINE H2 RECEPTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,485
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0159 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 236184
US-08-748-485-6

Query Match 1.1%; Score 8; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 VLTLLILI 501
Db 24 VLTLLILI 31

RESULT 11
US-08-208-007A-13
Sequence 13, Application US/08208007A
Patent No. 5501969
GENERAL INFORMATION:
APPLICANT: HASTINGS, ET AL.
TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/208,007A
FILING DATE: March 8, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 5501969e
FILING DATE: No. 5501969e
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-208-007A-13

Query Match 1.1%; Score 8; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLVL 12
Db 5 LLLLLLVL 12

RESULT 12
US-09-032-523-9
Sequence 9, Application US/09032523
Patent No. 6232454
GENERAL INFORMATION:
APPLICANT: Bandman, Olga

QY 5 LLLLLLVL 12
Db 5 LLLLLLVL 12

Query Match 1.1%; Score 8; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl
APPLICANT: Baugh, Mariah
TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0479 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 181994
US-09-032-523-9

Query Match 1.1%; Score 8; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLVL 12
Db 5 LLLLLLVL 12

RESULT 13
US-08-915-095A-13
Sequence 13, Application US/08915095A
Patent No. 6383793
GENERAL INFORMATION:
APPLICANT: Hastings, et al.
TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: PF107DA
CURRENT APPLICATION NUMBER: US/08/915,095A
CURRENT FILING DATE: 1997-08-20
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 396
TYPE: PRT
ORGANISM: Homo sapiens
US-08-915-095A-13

Query Match 1.1%; Score 8; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLVL 12
Db 5 LLLLLLVL 12

Query Match 1.1%; Score 8; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLVL 12
Db 5 LLLLLLVL 12

Query Match 1.1%; Score 8; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLL 12
| | | | |
Db 5 LLLLLVLL 12

RESULT 14

US-08-798-096-13
; Sequence 13, Application US/08798096
; Patent No. 6387682
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D2
; CURRENT APPLICATION NUMBER: US/08/798,096
; CURRENT FILING DATE: 1997-02-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-798-096-13

Query Match 1.1%; Score 8; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLL 12
| | | | |
Db 5 LLLLLVLL 12

RESULT 15

US-08-798-095A-13
; Sequence 13, Application US/08798095A
; Patent No. 6423507
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D3
; CURRENT APPLICATION NUMBER: US/08/798,095A
; CURRENT FILING DATE: 1997-02-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-798-095A-13

Query Match 1.1%; Score 8; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLVLL 12
| | | | |
Db 5 LLLLLVLL 12

Search completed: January 21, 2003, 09:58:21
Job time : 14.1324 secs

GenCore version 5.1.3
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:57:21 ; Search time 8.43994 Seconds
(without alignments)
1717.007 Million cell updates/sec

Title: US-10-060-830-3

Perfect score: 729

Sequence: 1 MPELLLLLLLLLEDDAGA.....TQEVSGAGRGDCDYFKEL 729

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 120991 seqs, 19878514 residues

Word size : 0

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications_AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	10.3	75	10	US-09-864-761-45371
2	10	1.4	677	10	US-09-006-298-21
3	10	1.4	696	10	US-09-925-300-1253
4	9	1.2	150	10	US-09-893-737-230
5	9	1.2	209	10	US-09-893-737-126
6	9	1.2	276	9	US-09-895-913A-356
7	9	1.2	1274	10	US-09-746-491-12
8	8	1.1	86	10	US-09-764-860-544
9	8	1.1	88	10	US-09-731-872-353
10	8	1.1	113	12	US-10-141-938-9
11	8	1.1	113	12	US-10-142-046-9
12	8	1.1	114	8	US-08-927-939-52
13	8	1.1	118	9	US-10-001-876-205
14	8	1.1	140	10	US-09-893-737-118
15	8	1.1	140	10	US-09-746-491-6
16	8	1.1	168	10	US-09-731-872-426
17	8	1.1	184	10	US-09-925-297-737
18	8	1.1	218	10	US-09-746-491-66
19	8	1.1	242	9	US-10-174-590-322

20	8	1.1	242	9	US-10-176-758-322	Sequence 322, App
21	8	1.1	242	10	US-09-746-491-68	Sequence 68, Appl
22	8	1.1	242	12	US-10-052-586-322	Sequence 322, App
23	8	1.1	283	10	US-09-925-299-984	Sequence 984, App
24	8	1.1	347	10	US-09-886-055-155	Sequence 155, App
25	8	1.1	396	10	US-09-215-450-22	Sequence 22, Appl
26	8	1.1	396	10	US-09-953-956-13	Sequence 13, Appl
27	8	1.1	396	12	US-10-114-464-13	Sequence 13, Appl
28	8	1.1	398	10	US-09-815-333-2	Sequence 2, Appl
29	8	1.1	398	12	US-09-842-316-2	Sequence 2, Appl
30	8	1.1	398	12	US-10-037-616-26	Sequence 26, Appl
31	8	1.1	400	9	US-09-971-228-12	Sequence 12, Appl
32	8	1.1	400	10	US-09-812-272-2	Sequence 2, Appl
33	8	1.1	400	12	US-10-037-616-27	Sequence 27, Appl
34	8	1.1	471	10	US-09-893-737-186	Sequence 186, App
35	8	1.1	480	9	US-10-084-018-8	Sequence 8, Appl
36	8	1.1	480	9	US-09-736-457-336	Sequence 336, App
37	8	1.1	480	9	US-09-902-941-336	Sequence 336, App
38	8	1.1	480	9	US-09-849-626-336	Sequence 336, App
39	8	1.1	523	10	US-09-930-218-4	Sequence 4, Appl
40	8	1.1	527	10	US-09-930-218-16	Sequence 16, Appl
41	8	1.1	673	10	US-09-782-980-68	Sequence 68, Appl
42	8	1.1	730	9	US-09-738-626-4196	Sequence 4196, Ap
43	8	1.1	764	9	US-10-045-624B-2	Sequence 2, Appl
44	8	1.1	829	10	US-09-905-983-50	Sequence 50, Appl
45	8	1.1	832	10	US-09-824-129-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-864-761-45371
; Sequence 45371, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 45371
LENGTH: 75
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC013497.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
OTHER INFORMATION: EST HUMAN HIT: A1565996.1, EVALUATE 2.00e-40
OTHER INFORMATION: SWISSPROT HIT: P98063, EVALUATE 1.00e-09
US-09-864-761-45371

Query Match 10.3%; Score 75; DB 10; Length 75;
Best Local Similarity 100.0%; Pred. No. 1.3e-60;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 DGGCHTVLGPESGLTSTINTPQTPNVCWEIRVMKGERVRIKFGDFIEDSDSCHFN 83
|||||
Db 1 DGGCHTVLGPESGLTSTINTPQTPNVCWEIRVMKGERVRIKFGDFIEDSDSCHFN 60

Qy 84 YLRIYNGIGVSRTEI 98
|||||
Db 61 YLRIYNGIGVSRTEI 75

RESULT 2
US-09-006-298-21
Sequence 21, Application US/09006298
Patent No. US20020082224A1
GENERAL INFORMATION:
APPLICANT: Jolly, Douglas J.
APPLICANT: Moore, Margaret D.
APPLICANT: Chada, Sunil
TITLE OF INVENTION: NON-IMMUNOGENIC PRODRUGS AND SELECTABLE
TITLE OF INVENTION: MARKERS FOR USE IN GENE THERAPY
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/006,298
FILING DATE: 13-JAN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McMasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 930049.459
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 677 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-006-298-21

Query Match 1.4%; Score 10; DB 10; Length 677;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLVLLLL 14
|||||
Db 9 LLLLLVLLLL 18

RESULT 3
US-09-925-300-1253
Sequence 1253, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben,
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1253
LENGTH: 696
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (5)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (541)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1253

Query Match 1.4%; Score 10; DB 10; Length 696;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLVLLLL 14
|||||
Db 28 LLLLLVLLLL 37

RESULT 4
US-09-893-737-230
Sequence 230, Application US/09893737
Patent No. US20020110855A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Presnell, Scott R.
TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
FILE REFERENCE: 00-41
CURRENT APPLICATION NUMBER: US/09/893,737
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,446
PRIOR FILING DATE: 2000-06-30
NUMBER OF SEQ ID NOS: 329
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 230
LENGTH: 150

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-230

Query Match 1.2%; Score 9; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LLLVLLLL 15
| | | | | | | |
Db 6 LLLVLLLL 14

RESULT 5
US-09-893-737-126
; Sequence 126, Application US/09893737
; Patent No. US20020110855A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 126
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-126

Query Match 1.2%; Score 9; DB 10; Length 209;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLVLLLL 13
| | | | | | | |
Db 9 LLLVLLLL 17

RESULT 6
US-09-895-913A-356
; Sequence 356, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1el Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 356
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-356

Query Match 1.2%; Score 9; DB 9; Length 276;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLVLLLL 13
| | | | | | | |
Db 16 LLLVLLLL 24

RESULT 7
US-09-746-491-12
; Sequence 12, Application US/09746491
; Patent No. US20020137202A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding:
; FILE REFERENCE: 15966-621
; CURRENT APPLICATION NUMBER: US/09/746,491
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: USSN 60/171,329
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1274
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-746-491-12

Query Match 1.2%; Score 9; DB 10; Length 1274;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLVLLLL 13
| | | | | | | |
Db 9 LLLVLLLL 17

RESULT 8
US-09-764-860-544
; Sequence 544, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 544
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (76)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-860-544

Query Match 1.1%; Score 8; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLVLLLL 12
| | | | | | | |
Db 9 LLLVLLLL 16

RESULT 9
US-09-731-872-353
; Sequence 353, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie

APPLICANT: Jobert, Severin
TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
FILE REFERENCE: 78.US3.REG
CURRENT APPLICATION NUMBER: US/09/731,872
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,629
PRIOR FILING DATE: 1999-12-08
PRIOR APPLICATION NUMBER: US 60/187,470
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 482
SOFTWARE: Patent.pm
SEQ ID NO 353
LENGTH: 88
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: -44...-1
US-09-731-872-353

Query Match 1.1%; Score 8; DB 10; Length 88;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PLFLLLLL 9
Db 32 PLFLLLLL 39

RESULT 10

US-10-141-938-9
Sequence 9, Application US/10141938
Patent No. US20020132305A1
GENERAL INFORMATION:

APPLICANT: Li et al.
TITLE OF INVENTION: Chemokine Alpha 3
FILE REFERENCE: PF251D2
CURRENT APPLICATION NUMBER: US/10/141,938
CURRENT FILING DATE: 2002-05-10
PRIOR APPLICATION NUMBER: 08/816,772
PRIOR FILING DATE: 1997-03-18
PRIOR APPLICATION NUMBER: 60/013,615
PRIOR FILING DATE: 1996-03-18
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 113
TYPE: PRT
ORGANISM: Homo sapiens

US-10-141-938-9

Query Match 1.1%; Score 8; DB 12; Length 113;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 LLVLLLLL 15
Db 21 LLVLLLLL 28

RESULT 11

US-10-142-046-9
Sequence 9, Application US/10142046
Patent No. US20020150994A1
GENERAL INFORMATION:

APPLICANT: Li et al.
TITLE OF INVENTION: Chemokine Alpha 3
FILE REFERENCE: PF251D3
CURRENT APPLICATION NUMBER: US/10/142,046
CURRENT FILING DATE: 2002-05-10
PRIOR APPLICATION NUMBER: 08/816,772
PRIOR FILING DATE: 1997-03-18
PRIOR APPLICATION NUMBER: 60/013,615

PRIOR FILING DATE: 1996-03-18
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 113
TYPE: PRT
ORGANISM: Homo sapiens
US-10-142-046-9

Query Match 1.1%; Score 8; DB 12; Length 113;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 LLVLLLLL 15
Db 21 LLVLLLLL 28

RESULT 12

US-08-927-939-52
Sequence 52, Application US/08927939
Patent No. US20010006640A1
GENERAL INFORMATION:

APPLICANT: Grainger, David J.
APPLICANT: Tatalick, Lauen Marie
TITLE OF INVENTION: Compounds and methods to inhibit or
TITLE OF INVENTION: augment an inflammatory response.
FILE REFERENCE: 295.022US1
CURRENT APPLICATION NUMBER: US/08/927,939
CURRENT FILING DATE: 1997-09-11
NUMBER OF SEQ ID NOS: 83
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 52
LENGTH: 114
TYPE: PRT
ORGANISM: Homo sapiens

US-08-927-939-52

Query Match 1.1%; Score 8; DB 8; Length 114;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 LLVLLLLL 15
Db 21 LLVLLLLL 28

RESULT 13

US-10-001-876-205
Sequence 205, Application US/10001876
Patent No. US20020177140A1
GENERAL INFORMATION:

APPLICANT: Salceda, Susana
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Herve
APPLICANT: Caffarkey, Robert
APPLICANT: Ali, Shujath
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua

TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes
FILE REFERENCE: DEX-0285
CURRENT APPLICATION NUMBER: US/10/001,876
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: 60/252,186
PRIOR FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 211
SOFTWARE: PatentIn version 3.1
SEQ ID NO 205
LENGTH: 118
TYPE: PRT
ORGANISM: Homo sapien

US-10-001-876-205

Query Match 1.1%; Score 8; DB 9; Length 118;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
 |||||
DB 27 LLVLLLLL 34

RESULT 14

US-09-893-737-118
; Sequence 118, Application US/09893737
; Patent No. US20020110855A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 118
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-118

Query Match 1.1%; Score 8; DB 10; Length 140;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
 |||||
DB 4 LLVLLLLL 11

RESULT 15

US-09-746-491-6
; Sequence 6, Application US/09746491
; Patent No. US20020137202A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-621
; CURRENT APPLICATION NUMBER: US/09/746,491
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: USSN 60/171,329
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-746-491-6

Query Match 1.1%; Score 8; DB 10; Length 140;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLVLLLLL 15
 |||||
DB 4 LLVLLLLL 11

Search completed: January 21, 2003, 10:00:38
Job time : 9.43994 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 12:56:47 ; Search time 1507.56 Seconds
(without alignments)
3222.856 Million cell updates/sec

Title: US-10-060-830-4
Perfect score: 300
Sequence: 1 aaacagaaagaaaaactg.....tatgcagacctagatccctta 300

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmd:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_othr:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	300	100.0	1065	13	BM450083
c	251	83.7	680	14	BQ014739
3	230	76.7	317	12	BE925427
c	58	19.3	357	12	BF511859
c	58	19.3	414	9	AI302412
6	30	10.0	912	12	BF167024
					BF167024 601774167

7	10.0	1272	11	AK006805
8	9.0	462	10	BB854159
9	8.7	421	9	AJ441540
10	8.7	628	14	BQ037529
11	8.7	689	9	AJ444823
12	8.7	764	9	AJ441638
13	8.7	809	9	AJ445175
c	7.3	578	17	B52631
14	7.3	817	9	AJ396784
15	7.0	505	17	BH339659
16	7.0	554	17	A0734458
c	7.0	672	12	BG536257
17	6.7	257	10	AW417733
c	6.7	514	17	A2055802
18	6.7	584	9	AL674742
19	6.7	725	17	AQ540114
c	6.7	740	17	AQ384330
20	6.7	803	13	B1558723
21	6.3	215	12	BE812059
22	6.3	255	12	BG467152
c	6.3	284	13	B1517143
23	6.3	298	9	AA114986
24	6.3	342	13	B1478206
c	6.3	360	12	BE777096
25	6.3	370	9	AI661435
c	6.3	378	9	AA286212
26	6.3	396	12	BF809441
27	6.3	427	9	AF114126
28	6.3	442	17	TAGH02P
c	6.3	468	13	B1511401
29	6.3	470	9	AI164382
30	6.3	470	12	BE869153
31	6.3	483	17	AZ495700
32	6.3	508	12	BE775551
33	6.3	516	9	AI463792
34	6.3	520	12	BE775912
c	6.3	524	12	BE775798
35	6.3	525	12	BE776396
c	6.3	528	14	BQ451882

ALIGNMENTS

RESULT 1
BM450083
LOCUS
DEFINITION
5', mRNA sequence.
AGENCOURT_6393434 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528102
BM450083
VERSION
BM450083.1 GI:18499123
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1065)
NTH-MGC http://mgc.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12204 row: g column: 15
High quality sequence stop: 665.
Location/Qualifiers
1. .1065

FEATURES
source

BASE COUNT	312 a	280 c	236 g	237 t	
ORIGIN	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5528102" /clone_lib="NIH_MGC_72" /tissue_type="melanotic melanoma" /lab_host="DH10B (phage-resistant)" /note="Organ: skin; vector: pCMV-SpORF6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."				
Query Match	100.0%	Score 300;	DB 13;	Length 1065;	
Best Local Similarity	100.0%	Pred. No. 1.1e-131;			
Matches 300;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY 1	AAACAGAAGAAAAAACTGAAGCACCATTGACTTACTCTTACTGGGACCGGCAGGTTG 60				
Db 213	AAACAGAAGAAAAAACTGAAGCACCATTGACTTACTCTTACTGGGACCGGCAGGTTG 272				
QY 61	GTGGAAGGAATGAAGCAGTTTCTTCCTGCAAAAGCAGTGGACCATGAGGAAACCCCACT 120				
Db 273	GTGGAAGGAATGAAGCAGTTTCTTCCTGCAAAAGCAGTGGACCATGAGGAAACCCCACT 332				
QY 121	TCGCTATAGCAGCAGCGAAGTTAATCACTGAGTCCAAAGAGAAGTCAACACAGTCTGCA 180				
Db 333	TCGCTATAGCAGCAGCGAAGTTAATCACTGAGTCCAAAGAGAAGTCAACACAGTCTGCA 392				
QY 181	GGCTGACTCTCCAGAGTATGCTCAGCCACTGGTAGGAGGAATGTTGTACATTCATCA 240				
Db 393	GGCTGACTCTCCAGAGTATGCTCAGCCACTGGTAGGAGGAATGTTGTACATTCATCA 452				
QY 241	AAGATCTACCTTTAAACCAGAAGAAAGAGGAGGAGGCTATCAGACCTAGATCCCTTA 300				
Db 453	AAGATCTACCTTTAAACCAGAAGAAAGAGGAGGAGGCTATCAGACCTAGATCCCTTA 512				
RESULT 2					
LOCUS	BQ014739/c				
DEFINITION	UI-H-ED1-axt-n12-0-UI.sl NCI_CGAP_ED1 Homo sapiens cDNA clone				
ACCESSION	IMAGE:5833547 3', mRNA sequence.				
VERSION	BQ014739				
KEYWORDS	BQ014739 1 GI:19739640				
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapsb@mail.nih.gov Tissue Procurement: Dr. Jose Mercuende cDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Dr. M. Bento Soares, University of Iowa Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Seq primer: M13 FORWARD POLY-A-Yes.				
FEATURES	Location/Qualifiers				
source	1. .680				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:5833547"				
	/clone_lib="NCI_CGAP_ED1"				
	/tissue_type="Chondrosarcoma"				
	/dev_stage="Adult"				

/lab_host="DH10B (Life Technologies)"									
/note=Organ: Left Pubic Bone; Vector: pT7T3-Pac									
/pharmac: with a modified polylinker; Site 1: Ecor I;									
Site 2: Not I; NCI CGAP EDI is a normalized cDNA library									
containing the following tissue(s): Chondrosarcoma cell									
line CS5. The library was constructed according to Bonaldi									
, Lennon and Soares, Genome Research, 6:791-806, 1996.									
First strand cDNA synthesis was primed with an oligo-dT									
primer containing a Not I site. Double stranded cDNA was									
ligated to an Ecor I adaptor, digested with Not I, and									
cloned directionally into pT7T3-Pac vector. The									
oligonucleotide used to prime the synthesis of									
first-strand cDNA contains a library tag sequence that is									
located between the Not I site and the (dT)18 tail. The									
sequence tag for this library is GCTCAAGGCT.									
TAG LIB=UI-H-EDI									
TAG_TISSUE=chondrosarcoma									
TAG_SEQ=CGTCAAGCT*									
BASE COUNT	138 a	157 c	172 g	213 t					
ORIGIN									
Query Match	83.7%	Score 251;	DB 14;	Length 680;					
Best Local Similarity	100.0%;	Pred. No. 2e-108;							
Matches 251;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
Qy 50	CGGCAGGTTGGTGGAAAGGAATGAAGCAGTTCCTTCCTGCAAAAGCAGTGGACCATGAG	109							
Db 680	CGGCAGGTTGGTGGAAAGGAATGAAGCAGTTCCTTCCTGCAAAAGCAGTGGACCATGAG	621							
Qy 110	GAACCCGAGTTCGCTATAGCAGCAGCGAAGTTAATCACCTGAGTCCACAGAGAAGTCACC	169							
Db 620	GAACCCGAGTTCGCTATAGCAGCAGCGAAGTTAATCACCTGAGTCCACAGAGAAGTCACC	561							
Qy 170	ACAGTGTGCGAGGCTGACTCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATGTGGT	229							
Db 560	ACAGTGTGCGAGGCTGACTCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATGTGGT	501							
Qy 230	ACATTTCATCAAGATCTACCTTTAAACAGAGAAGGAAAGAACGAGCGCTATGCAGAC	289							
Db 500	ACATTTCATCAAGATCTACCTTTAAACAGAGAAGGAAAGAACGAGCGCTATGCAGAC	441							
Qy 290	CTAGATCCTTA 300								
Db 440	CTAGATCCTTA 430								
RESULT 3									
BE925427									
LOCUS	BE925427	317 bp	mRNA	linear	EST 02-OCT-2000				
DEFINITION	PM0-AN0087-300800-002-cl1 AN0087 Homo sapiens cDNA, mRNA sequence.								
ACCESSION	BE925427								
VERSION	BE925427.1	GI:10451503							
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
	1 (bases 1 to 317)								
	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,								
	Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,								
	Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,								
	Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare								
	M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and								
	Simpson, A.J.								
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed								
	sequence tags								
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)								
MEDLINE	2002663								
COMMENT	Contact: Simpson A.J.G.								
	Laboratory of Cancer Genetics								
	Ludwig Institute for Cancer Research								
	Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,								
	Brazil								

Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-PMO-AN0087-300800-002-cl1&t3=2000-08-30&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 13
 High quality sequence stop: 316.

FEATURES

Location/Qualifiers
 1. 317
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="AN0087"
 /dev_stage="Adult"
 /note="Organ: amnion_normal; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions." 75 t

BASE COUNT 93 a 69 c 80 g 75 t
 ORIGIN
 Query Match 76.7%; Score 230; DB 12; Length 317;
 Best Local Similarity 100.0%; Pred. No. 2.1e-98;
 Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 ACTTACTTCTGCGGACCGGAGGTTGGTGAAGGATGAGCAGGTTCTTCTCGCAA 92
 |||||
 Db 88 ACTTACTTCTGCGGACCGGAGGTTGGTGAAGGATGAGCAGGTTCTTCTCGCAA 147
 |||||
 QY 93 AGCAGTGGACCATGAGGAACCCAGTTCGGCTATAGCAGCAGCGAAGTTAATCACTGA 152
 |||||
 Db 148 AGCAGTGGACCATGAGGAACCCAGTTCGGCTATAGCAGCAGCGAAGTTAATCACTGA 207
 |||||
 QY 153 GTCGAAGAAGTCACACAGTCTGCAGGCTGACCTGACGATGCTGCAGCCACTGG 212
 |||||
 Db 208 GTCGAAGAAGTCACACAGTCTGCAGGCTGACCTGACGATGCTGCAGCCACTGG 267
 |||||
 QY 213 TAGGAGGAATGTTGGTACACTTCATCAAGATCTACCTTTAAACAGAA 262
 |||||
 Db 268 TAGGAGGAATGTTGGTACACTTCATCAAGATCTACCTTTAAACAGAA 317
 |||||

RESULT 4
 BF511859/c
 LOCUS
 DEFINITION UI-B14-aps-e-06-0-UI-s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
 IMAGE:3088594 3', mRNA sequence.
 ACCESSION BF511859
 VERSION BF511859.1 GI:11595157
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 357)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov

REFERENCE 1 (bases 1 to 357)
 AUTHORS NCI-CGAP
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
 NCI-CGAP clone distribution information can be found through the
 I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

Location/Qualifiers
 1. 357
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:3088594"
 /clone_lib="NCI_CGAP_Sub8"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub8
 is a subtracted library derived from NCI_CGAP_Sub5. The
 NCI_CGAP_Sub8 library had 2.5 million recombinants. A
 single-stranded DNA preparation of NCI_CGAP_Sub5 was used
 as a tracer in a subtractive hybridization with a driver
 comprising a pool of clones from NCI_CGAP_Sub5 (IMAGE
 clone ids 2732833-2737415, 3068040-3069191; 25% of the
 driver population), a pool of clones from NCI_CGAP_Sub4
 (IMAGE clone ids 2723592-2729326; 25% of the driver
 population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE Ids
 2728969-2733190; 25% of the driver population), and
 NCI_CGAP_Sub7 (IMAGE Ids 3069192-3072238, 3081864-3084550
 ; 25% of the driver population). Subtraction was
 performed as previously described [Bonaldi, Lennon &
 Soares (1996): Normalization and Subtraction: Two
 Approaches To Facilitate Gene Discovery. Genome Research
 6, 791-806.
 TAG_SEQ=None found"
 99 a 69 c 82 g 107 t

BASE COUNT 99 a 69 c 82 g 107 t
 ORIGIN
 Query Match 19.3%; Score 58; DB 12; Length 357;
 Best Local Similarity 100.0%; Pred. No. 6.4e-17;
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACAGAAAGAAAAAAGTGAAGGACCTATGACTTACTTGGACCGGCAGGT 58
 |||||
 Db 283 AAACAGAAAGAAAAAAGTGAAGGACCTATGACTTACTTGGACCGGCAGGT 226
 |||||

RESULT 5
 AI302412/c
 LOCUS
 DEFINITION qn51a10.x1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1901754 3',
 mRNA sequence.
 ACCESSION AI302412
 VERSION AI302412.1 GI:3961758
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 414)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 778 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 379.

FEATURES
 Location/Qualifiers
 1. 414
 /organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:1901754"
/clone_lib="NCI_CGAP_Kid5"
/tissue_type="2_pooled tumors (clear cell type)"
/lab_host="DH10B"
/notes="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
AACTGGAAGAAATTCGGCGCGCAATATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 113 a 81 c 99 g 121 t
ORIGIN

Query Match 19.3%; Score 58; DB 9; Length 414;
Best Local Similarity 100.0%; Pred. No. 6.2e-17;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACAGAAAGAAAAAAGCAGCCTATGACTTACCTTACTTGGGACCGGCAGGT 58
|||||
DB 275 AAACAGAAAGAAAAAAGCAGCCTATGACTTACCTTACTTGGGACCGGCAGGT 218
|||||

RESULT 6
LOCUS BF167024 912 bp mRNA linear EST 30-OCT-2000
DEFINITION 601774167F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3994745 5',
mRNA sequence.
ACCESSION BF167024
VERSION BF167024.1 GI:11047286
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 912)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9213 row: e column: 18
High quality sequence stop: 651.
FEATURES
source Location/Qualifiers
1..912
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3994745"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 233 a 265 c 256 g 158 t
ORIGIN

Query Match 10.0%; Score 30; DB 12; Length 912;
Best Local Similarity 100.0%; Pred. No. 0.00096;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 50 CGGCAGCTTGGTGAAGGATGAAGCAG 79
|||||
DB 172 CGGCAGCTTGGTGAAGGATGAAGCAG 201
|||||

RESULT 7
LOCUS AK006805 1272 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:1700055p21:endothelial and smooth muscle
cell-derived neuropilin-like molecule, full insert sequence.
ACCESSION AK006805
VERSION AK006805.1 GI:12840068
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
clone_lib:RIKEN full-length enriched mouse cDNA library
clone:1700055p21.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaide,I., Pesole,G.,
Quackenbush,J., Schriml,L.M., Staabli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,F., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toto-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohsaki,S.
and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5 (bases 1 to 1272)

```


COMMENT On Mar 27, 2002 this sequence version replaced gi:19771069.
Contact: Joan Burnside

```

LOCUS      AJ441638      764 bp      mRNA      linear      EST 19-APR-2002
DEFINITION AJ441638      dkfz426 Gallus gallus cDNA clone 128f4r1, mRNA sequence.
ACCESSION  AJ441638
VERSION    AJ441638.1  GI:20208859
KEYWORDS   EST.
SOURCE     chicken.
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 764)
AUTHORS   Buerstedde,J.M.
TITLE     Gallus gallus bursal lymphocyte EST
JOURNAL   Unpublished (2002)
COMMENT   Contact: Buerstedde JM
            Cellular Immunology
            Heinrich-Pette-Institut
            Martinistr. 52, 20251 Hamburg, Germany
            Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES   Location/Qualifiers
            source
            1..764
            /organism="Gallus gallus"
            /strain="CB"
            /db_xref="taxon:9031"
            /clone="128f4r1"
            /clone_lib="dkfz426"
            /tissue_type="Bursa of Fabricius"
            /cell_type="bursal lymphocyte"
            /dev_stage="2-3 weeks old"
BASE COUNT 222 a 186 c 175 g 181 t
ORIGIN
Query Match      8.7%; Score 26; DB 9; Length 764;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 59 TGGTGGAAAGGAATGAAGCAGTTTCT 84
|||||
DB 54 TGGTGGAAAGGAATGAAGCAGTTTCT 79
|||||

RESULT 13
AJ445175
LOCUS      AJ445175      809 bp      mRNA      linear      EST 19-APR-2002
DEFINITION AJ445175      dkfz426 Gallus gallus cDNA clone 92n5r1, mRNA sequence.
ACCESSION  AJ445175
VERSION    AJ445175.1  GI:20212396
KEYWORDS   EST.
SOURCE     chicken.
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 809)
AUTHORS   Buerstedde,J.M.
TITLE     Gallus gallus bursal lymphocyte EST
JOURNAL   Unpublished (2002)
COMMENT   Contact: Buerstedde JM
            Cellular Immunology
            Heinrich-Pette-Institut
            Martinistr. 52, 20251 Hamburg, Germany
            Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES   Location/Qualifiers
            source
            1..809
            /organism="Gallus gallus"
            /strain="CB"
            /db_xref="taxon:9031"
            /clone="92n5r1"
            /clone_lib="dkfz426"
            /tissue_type="Bursa of Fabricius"
            /cell_type="bursal lymphocyte"
            /dev_stage="2-3 weeks old"
BASE COUNT 237 a 192 c 180 g 198 t 2 others

```

```

ORIGIN
Query Match      8.7%; Score 26; DB 9; Length 809;
Best Local Similarity 100.0%; Pred. No. 0.077;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 59 TGGTGGAAAGGAATGAAGCAGTTTCT 84
|||||
DB 2  TGGTGGAAAGGAATGAAGCAGTTTCT 27
|||||

RESULT 14
B52631/c
LOCUS      B52631      578 bp      DNA      linear      GSS 20-JUN-1998
DEFINITION CIT-HSP-2006B7.TR CIT-HSP Homo sapiens genomic clone 2006B7, DNA
            sequence.
ACCESSION  B52631
VERSION    B52631.1  GI:2606965
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 578)
AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
            and Venter,J.C.
TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
JOURNAL    Unpublished (1997)
COMMENT    Other GSSs: CIT-HSP-2006B7.TFB
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES   Location/Qualifiers
            source
            1..578
            /organism="Homo sapiens"
            /db_xref="GDB:7039521"
            /db_xref="taxon:9606"
            /clone="2006B7"
            /clone_lib="CIT-HSP"
            /sex="Male"
            /cell_type="Sperm"
            /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
            HindIII"
BASE COUNT 171 a 113 c 115 g 179 t
ORIGIN
Query Match      7.7%; Score 23; DB 17; Length 578;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 94 AGCAGTGGACCATGAGGAACCC 116
|||||
DB 181 AGCAGTGGACCATGAGGAACCC 159
|||||

RESULT 15
AJ396784
LOCUS      AJ396784      817 bp      mRNA      linear      EST 25-JAN-2001
DEFINITION AJ396784      dkfz426 Gallus gallus cDNA clone 27m24r1, mRNA sequence.
ACCESSION  AJ396784
VERSION    AJ396784.1  GI:7129070
KEYWORDS   EST.

```

SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 817)
AUTHORS Abdrakhmanov, I., Lodygin, D., Geroth, P., Arakawa, H., Law, A., Plachy
J., Korn, B. and Buerstedde, J.M.
TITLE A large database of chicken bursal ESTs as a resource for the
analysis of vertebrate gene function
JOURNAL Genome Res. 10 (12), 2062-2069 (2000)
MEDLINE 20568495
COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.
Location/Qualifiers
1. .817
/organism="Gallus gallus"
/strain="CB"
/db_xref="taxon:9031"
/clone="27m24r1"
/clone_lib="dkfz426"
/tissue_type="Bursa of Fabricius"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
BASE COUNT 240 a 194 c 178 g 203 t 2 others
ORIGIN
Query Match 7.3%; Score 22; DB 9; Length 817;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 63 GGAAGGAGTGAAGCAGTTTCT 84
|||||
Db 1 GGAAGGAGTGAAGCAGTTTCT 22

Search completed: January 21, 2003, 14:43:19
Job time : 1512.56 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 12:59:12 ; Search time 38.2677 Seconds
(without alignments)
2404.194 Million cell updates/sec

Title: US-10-060-830-4
Perfect score: 300
Sequence: 1 aacagagaagaaaaaactg.....tatgcagacctagatcctta 300

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/FCUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	183	61.0	640	US-09-328-111-480	Sequence 480, App
c 2	20	6.7	11091	US-09-134-001C-2243	Sequence 2243, App
c 3	19	6.3	4129	US-08-370-319C-12	Sequence 12, App
c 4	19	6.3	4129	US-09-224-834-12	Sequence 12, App
c 5	17	5.7	80	US-08-718-904-110	Sequence 110, App
c 6	17	5.7	267	US-08-931-858E-84	Sequence 84, App
c 7	17	5.7	267	US-08-931-858E-85	Sequence 85, App
c 8	17	5.7	267	US-08-981-739-84	Sequence 84, App
c 9	17	5.7	267	US-08-981-739-85	Sequence 85, App
c 10	17	5.7	267	US-09-128-026-84	Sequence 84, App
c 11	17	5.7	267	US-09-128-026-85	Sequence 85, App
c 12	17	5.7	273	US-08-931-858E-86	Sequence 86, App
c 13	17	5.7	273	US-08-981-739-86	Sequence 86, App
c 14	17	5.7	273	US-09-128-026-86	Sequence 86, App
c 15	17	5.7	291	US-08-931-858E-183	Sequence 183, App
c 16	17	5.7	291	US-08-931-858E-184	Sequence 184, App
c 17	17	5.7	291	US-08-931-858E-194	Sequence 194, App
c 18	17	5.7	291	US-08-931-858E-195	Sequence 195, App
c 19	17	5.7	291	US-08-981-739-174	Sequence 174, App
c 20	17	5.7	291	US-08-981-739-176	Sequence 176, App
c 21	17	5.7	291	US-09-128-026-174	Sequence 174, App
c 22	17	5.7	291	US-09-128-026-176	Sequence 176, App
c 23	17	5.7	339	US-09-188-930-93	Sequence 93, App
c 24	17	5.7	339	US-09-188-930-267	Sequence 267, App
c 25	17	5.7	391	US-08-931-858E-107	Sequence 107, App
c 26	17	5.7	391	US-08-981-739-107	Sequence 107, App
c 27	17	5.7	391	US-09-128-026-107	Sequence 107, App

c 28	17	5.7	405	4	US-08-981-739-175	Sequence 175, App
c 29	17	5.7	405	4	US-09-128-026-175	Sequence 175, App
c 30	17	5.7	462	3	US-08-718-904-118	Sequence 118, App
c 31	17	5.7	471	4	US-08-931-858E-179	Sequence 179, App
c 32	17	5.7	471	4	US-08-931-858E-180	Sequence 180, App
c 33	17	5.7	471	4	US-08-931-858E-190	Sequence 190, App
c 34	17	5.7	471	4	US-08-931-858E-191	Sequence 191, App
c 35	17	5.7	486	4	US-09-358-972-90	Sequence 90, App
c 36	17	5.7	486	4	US-09-406-065-78	Sequence 78, App
c 37	17	5.7	526	1	US-08-686-878A-35	Sequence 35, App
c 38	17	5.7	526	4	US-09-175-928-35	Sequence 35, App
c 39	17	5.7	544	4	US-08-931-858E-105	Sequence 105, App
c 40	17	5.7	544	4	US-08-981-739-105	Sequence 105, App
c 41	17	5.7	544	4	US-09-128-026-105	Sequence 105, App
c 42	17	5.7	559	4	US-08-931-858E-131	Sequence 131, App
c 43	17	5.7	559	4	US-08-931-858E-188	Sequence 188, App
c 44	17	5.7	559	4	US-08-931-858E-189	Sequence 189, App
c 45	17	5.7	559	4	US-08-981-739-131	Sequence 131, App

ALIGNMENTS

RESULT 1
US-09-328-111-480/c
; Sequence 480, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328.111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088.801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-480

Query Match	61.0%	Score 183;	DB 4;	Length 640;
Best Local Similarity	100.0%;	Pred. No. 4.4e-80;		
Matches 183;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	49	CCGGCAGGTTGGTGGAAAGGAATGAACAGTTTCTTCTCTGCAAAAGCAGTGGACCATGA	108	
Db	184	CCGGCAGGTTGGTGGAAAGGAATGAACAGTTTCTTCTGCAAAAGCAGTGGACCATGA	125	
QY	109	GGAACCCCACTTCCTATAGCAGCAGGCAAGTTAATCACCTGAGTCCCAAGAGAAGTCAC	168	
Db	124	GGAACCCCACTTCCTATAGCAGCAGGCAAGTTAATCACCTGAGTCCCAAGAGAAGTCAC	65	
QY	169	CACAGTGTGAGGCTGACTCTGCAGAGTATGCTCAGGCACCTGGTAGGAGGAATTTGG	228	

Db 64 CACAGTCTGAGGCTGACTCTGCAGAGATGCTCAGCCACTGGTAGGAGGAATTGTTGG 5

QY 229 TAC 231
 |||
 Db 4 TAC 2

RESULT 2

US-09-134-001C-2243/c
 ; Sequence 2243, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US/09/134,001C
 ; PRIOR FILING DATE: 1997-11-08 US 60/064,964
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 2243
 ; LENGTH: 11091
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-2243

Query Match 6.7%; Score 20; DB 4; Length 11091;
 Best Local Similarity 100.0%; Pred. No. 0.81;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 CTATAGCAGCGCAAGTTA 143
 |||
 Db 1300 CTATAGCAGCGCAAGTTA 1281

RESULT 3

US-08-370-319C-12/c
 ; Sequence 12, Application US/08370319C
 ; Patent No. 5856091
 ; GENERAL INFORMATION:
 ; APPLICANT: Brichard, Vincent; Van Pel, Aline;
 ; APPLICANT: Traversari, Catia; W lfel, Thomas; Coullie, Pierre;
 ; APPLICANT: Boon-Falleur, Thierry; De plaen, Etienne
 ; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
 ; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PERCURSOR PROCESSED TO AT LEAST ONE TU
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felfe & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/370,319C
 ; FILING DATE: 10-JANUARY-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/272,351
 ; FILING DATE: 8-JULY-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/032,978
 ; FILING DATE: 18-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, No. 5856091man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5377.1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 688-9200
 ; TELEFAX: (212) 838-3884
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4129 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; FEATURE:
 ; OTHER INFORMATION: The sequence is preceded by an
 ; OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
 ; OTHER INFORMATION: kilobases
 ; US-08-370-319C-12

Query Match 6.3%; Score 19; DB 2; Length 4129;
 Best Local Similarity 100.0%; Pred. No. 2.6;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AACAGAAAGAAAAAACTG 20
 |||
 Db 1112 AACAGAAAGAAAAAACTG 1094

RESULT 4

US-09-224-834-12/c
 ; Sequence 12, Application US/09224834
 ; Patent No. 6201111
 ; GENERAL INFORMATION:
 ; APPLICANT: Brichard, Vincent; Van Pel, Aline;
 ; APPLICANT: Traversari, Catia; W lfel, Thomas; Coullie, Pierre;
 ; APPLICANT: Boon-Falleur, Thierry; De plaen, Etienne
 ; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
 ; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PERCURSOR PROCESSED TO AT LEAST O
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felfe & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/224,834
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/370,319
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/032,978
 ; FILING DATE: 18-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hanson, No. 6201111man D.
 ; REGISTRATION NUMBER: 30,946
 ; REFERENCE/DOCKET NUMBER: LUD 5377.1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 688-9200
 ; TELEFAX: (212) 838-3884
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4129 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double

APPLICATION NUMBER: US/08/931,858E
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 971486
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:
LENGTH: 267 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-931-858E-85

Query Match 5.7%; Score 17; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 GCTCAGCCACTGGTAGG 216
|||||
Db 37 GCTCAGCCACTGGTAGG 21

RESULT 8

US-08-981-739-84/c
Sequence 84, Application US/08981739
Patent No. 6232449
GENERAL INFORMATION:

APPLICANT: JOHNSON JR., EUGENE M.
MILBRANDT, JEFFREY D.
KOTZBAUER, PAUL T.
LAMPE, PATRICIA A.

TITLE OF INVENTION: PERSEPHIN AND RELATED GROWTH FACTORS
NUMBER OF SEQUENCES: 176
CORRESPONDENCE ADDRESS:

ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: US

ZIP: 63105-1817

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/981,739

FILING DATE: 31-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US97/03461

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: HOLLAND, DONALD R.

REGISTRATION NUMBER: 35,197

REFERENCE/DOCKET NUMBER: 976163

TELEPHONE: 314-727-5188

TELEFAX: 314-727-6092

INFORMATION FOR SEQ ID NO: 84:

SEQUENCE CHARACTERISTICS:

LENGTH: 267 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cdna

SEQUENCE DESCRIPTION: SEQ ID NO: 84:

US-08-981-739-84

Query Match 5.7%; Score 17; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 GCTCAGCCACTGGTAGG 216
|||||
Db 37 GCTCAGCCACTGGTAGG 21

RESULT 9

US-08-981-739-85/c
Sequence 85, Application US/08981739
Patent No. 6232449
GENERAL INFORMATION:

APPLICANT: JOHNSON JR., EUGENE M.
MILBRANDT, JEFFREY D.
KOTZBAUER, PAUL T.
LAMPE, PATRICIA A.

TITLE OF INVENTION: PERSEPHIN AND RELATED GROWTH FACTORS
NUMBER OF SEQUENCES: 176
CORRESPONDENCE ADDRESS:

ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
CITY: ST. LOUIS
STATE: MISSOURI
COUNTRY: US

ZIP: 63105-1817

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/981,739

FILING DATE: 31-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US97/03461

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: HOLLAND, DONALD R.

REGISTRATION NUMBER: 35,197

REFERENCE/DOCKET NUMBER: 976163

TELEPHONE: 314-727-5188

TELEFAX: 314-727-6092

INFORMATION FOR SEQ ID NO: 85:

SEQUENCE CHARACTERISTICS:

LENGTH: 267 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cdna

SEQUENCE DESCRIPTION: SEQ ID NO: 85:

US-08-981-739-85

Query Match 5.7%; Score 17; DB 4; Length 267;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 GCTCAGCCACTGGTAGG 216
|||||
Db 37 GCTCAGCCACTGGTAGG 21

RESULT 10

US-09-128-026-84/c
Sequence 84, Application US/09128026
Patent No. 6403335
GENERAL INFORMATION:

APPLICANT: JOHNSON JR., EUGENE M.

```
; APPLICANT: MILBRANDT, JEFFREY D.
; APPLICANT: KOTZBAUER, PAUL T.
; APPLICANT: LAMPE, PATRICIA A.
; TITLE OF INVENTION: PERSEPHIN AND RELATED GROWTH FACTORS
; NUMBER OF SEQUENCES: 176
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: US
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,026
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 976163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 267 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-128-026-84
;
; Query Match 5.7%; Score 17; DB 4; Length 267;
; Best Local Similarity 100.0%; Pred. No. 26;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 200 GCTCAGCCACTGGTAGG 216
Db 37 GCTCAGCCACTGGTAGG 21
|||||
;
; RESULT 11
; US-09-128-026-85/c
; Sequence 85, Application US/09128026
; Patent No. 6403335
; GENERAL INFORMATION:
; APPLICANT: JOHNSON JR., EUGENE M.
; APPLICANT: MILBRANDT, JEFFREY D.
; APPLICANT: KOTZBAUER, PAUL T.
; APPLICANT: LAMPE, PATRICIA A.
; TITLE OF INVENTION: PERSEPHIN AND RELATED GROWTH FACTORS
; NUMBER OF SEQUENCES: 176
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: US
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,026
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 976163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 267 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-128-026-84
;
; Query Match 5.7%; Score 17; DB 4; Length 267;
; Best Local Similarity 100.0%; Pred. No. 26;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 200 GCTCAGCCACTGGTAGG 216
Db 37 GCTCAGCCACTGGTAGG 21
|||||
;
; RESULT 12
; US-08-931-858E-86/c
; Sequence 86, Application US/08931858E
; Patent No. 6222022
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, EUGENE M.
; APPLICANT: MILBRANDT, JEFFREY D.
; APPLICANT: KOTZBAUER, PAUL T.
; APPLICANT: LAMPE, PATRICIA A.
; APPLICANT: KLEIN, ROBERT
; APPLICANT: DESAUVAGE, FRED
; TITLE OF INVENTION: PERSEPHIN AND RELATED GROWTH FACTOR
; NUMBER OF SEQUENCES: 239
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,858E
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 971486
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-931-858E-86
;
; Query Match 5.7%; Score 17; DB 4; Length 273;
; Best Local Similarity 100.0%; Pred. No. 26;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 976163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 267 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-128-026-85
;
; Query Match 5.7%; Score 17; DB 4; Length 267;
; Best Local Similarity 100.0%; Pred. No. 26;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 200 GCTCAGCCACTGGTAGG 216
Db 37 GCTCAGCCACTGGTAGG 21
|||||
;
; RESULT 12
; US-08-931-858E-86/c
; Sequence 86, Application US/08931858E
; Patent No. 6222022
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, EUGENE M.
; APPLICANT: MILBRANDT, JEFFREY D.
; APPLICANT: KOTZBAUER, PAUL T.
; APPLICANT: LAMPE, PATRICIA A.
; APPLICANT: KLEIN, ROBERT
; APPLICANT: DESAUVAGE, FRED
; TITLE OF INVENTION: PERSEPHIN AND RELATED GROWTH FACTOR
; NUMBER OF SEQUENCES: 239
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,858E
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 971486
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-931-858E-86
;
; Query Match 5.7%; Score 17; DB 4; Length 273;
; Best Local Similarity 100.0%; Pred. No. 26;
; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 200 GCTCAGCCACTGGTAGG 216
Db 37 GCTCAGCCACTGGTAGG 21

RESULT 13

US-08-981-739-86/C
; Sequence 86, Application US/08981739
; Patent No. 6232449
; GENERAL INFORMATION:
; APPLICANT: JOHNSON JR., EUGENE M.
; MILBRANDT, JEFFREY D.
; KOTZBAUER, PAUL T.
; LAMPE, PATRICIA A.
; TITLE OF INVENTION: PERSEPHIN AND RELATED GROWTH FACTORS
; NUMBER OF SEQUENCES: 176
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: US
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/981,739
; FILING DATE: 31-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/03461
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 976163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-08-981-739-86

Query Match 5.7%; Score 17; DB 4; Length 273;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 GCTCAGCCACTGGTAGG 216
Db 37 GCTCAGCCACTGGTAGG 21

RESULT 14

US-09-128-026-86/C
; Sequence 86, Application US/09128026
; Patent No. 6403335
; GENERAL INFORMATION:
; APPLICANT: JOHNSON JR., EUGENE M.
; MILBRANDT, JEFFREY D.
; KOTZBAUER, PAUL T.
; LAMPE, PATRICIA A.
; TITLE OF INVENTION: PERSEPHIN AND RELATED GROWTH FACTORS
; NUMBER OF SEQUENCES: 176

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MISSOURI
; COUNTRY: US
; ZIP: 63105-1817
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,026
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 976163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 727-5188
; TELEFAX: (314) 727-6092
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-128-026-86

Query Match 5.7%; Score 17; DB 4; Length 273;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 GCTCAGCCACTGGTAGG 216
Db 37 GCTCAGCCACTGGTAGG 21

RESULT 15

US-08-931-858E-183/C
; Sequence 183, Application US/08931858E
; Patent No. 6222022
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, EUGENE M
; MILBRANDT, JEFFREY D
; KOTZBAUER, PAUL T
; LAMPE, PATRICIA A
; KLEIN, ROBERT
; DESAUVAGE, FRED
; TITLE OF INVENTION: PERSEPHIN AND RELATED GROWTH FACTOR
; NUMBER OF SEQUENCES: 239
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BOULEVARD, SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,858E
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197

; REFERENCE/DOCKET NUMBER: 971486
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 183:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-931-858E-183

Query Match 5.7%; Score 17; DB 4; Length 291;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 GCTCAGCCACTGGTAGG 216
Db 52 GCTCAGCCACTGGTAGG 36

Search completed: January 21, 2003, 14:44:54
Job time : 42.2677 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	300	100.0	0	5657	9	US-09-974-238-96	Sequence 96, Appl
c	183	61.0	640	10	US-09-879-536-480		Sequence 480, App
	109	36.3	580	10	US-09-864-761-12400		Sequence 12400, A
	106	35.3	467	10	US-09-864-761-30344		Sequence 30344, A
4	19	6.3	423	10	US-09-864-761-18355		Sequence 18355, A
c	6	19	6.3	7105	10	US-09-864-777-3815	Sequence 3815, Ap
	7	19	6.3	30365	10	US-09-825-414-1	Sequence 1, Appl1
8	18	6.0	370	10	US-09-867-701-490		Sequence 490, App
9	18	6.0	452	10	US-09-770-444-644		Sequence 644, App
10	18	6.0	581	10	US-09-864-761-14320		Sequence 14320, A
11	18	6.0	735	9	US-09-938-8424-1349		Sequence 1349, Ap
c	12	18	6.0	789	10	US-09-070-927A-506	Sequence 506, App
c	13	17	5.7	123	10	US-09-864-761-27614	Sequence 27614, A
c	14	17	5.7	171	10	US-09-878-574-8457	Sequence 8457, Ap
c	15	17	5.7	224	10	US-09-864-761-30968	Sequence 30968, A
c	16	17	5.7	396	10	US-09-867-701-3900	Sequence 3900, Ap
c	17	17	5.7	449	10	US-09-864-761-10977	Sequence 10977, A
c	18	17	5.7	456	10	US-09-770-444-533	Sequence 533, App
c	19	17	5.7	469	10	US-09-864-761-14412	Sequence 14412, A

QY 181 GGCTGACTGTCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTACACTTCATCA 240
|||||
Db 1707 GGCTGACTGTCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTACACTTCATCA 1766
|||||
QY 241 AGAGTCTACTTTAAACCAGAGAAAGAAAGCAGGCTATGCGAGACCTAGATCTCTTA 300
|||||
Db 1767 AAGATCTACTCTTTAAACCAGAGAAAGAAAGCAGGCTATGCGAGACCTAGATCTCTTA 1826
|||||
RESULT 2
US-09-879-536-480/c
; Sequence 480, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/879, 536
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/088, 801
; PRIOR FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Query Match 61.08; Score 183; DB 10; Length 640;
Best Local Similarity 100.0%; Pred. No. 1.7e-82;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 CCGGCGAGTTGGTGGAAAGGAATGAAGCAGTTTCTCTGCAAAAGCAGTGGACCATGA 108
|||||
Db 184 CCGGCGAGTTGGTGGAAAGGAATGAAGCAGTTTCTCTGCAAAAGCAGTGGACCATGA 125
|||||
QY 109 GGAACCCCGAGTTGCGCTATAGCAGCAGCAAGTTTAATCACCTGAGTCCAGAGAGTCCAC 168
|||||
Db 124 GGAACCCCGAGTTGCGCTATAGCAGCAGCAAGTTTAATCACCTGAGTCCAGAGAGTCCAC 65
|||||
QY 169 CACAGTGTGTCAGGCTGACTCTGTCAGAGTATGCTCAGCCACTGTAGAGGAATTTGG 228
|||||
Db 64 CACAGTGTGTCAGGCTGACTCTGTCAGAGTATGCTCAGCCACTGTAGAGGAATTTGG 5
|||||
QY 229 TAC 231
|||
Db 4 TAC 2

RESULT 3
US-09-864-761-12400
; Sequence 12400, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Weusheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers...1.1
; SEQ ID NO 12400
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
US-09-864-761-12400

Query Match 36.3%; Score 109; DB 10; Length 580;
Best Local Similarity 100.0%; Pred. No. 2.2e-45;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 CAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTACACTTCATCAAGATCTACCT 251
|||||
Db 68 CAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTACACTTCATCAAGATCTACCT 127
|||||
QY 252 TTAACCCAGAGAGAGAAAGAACGAGGCTATGCGAGACCTAGATCCCTTA 300
|||||
Db 128 TTAACCCAGAGAGAGAGAAAGAACGAGGCTATGCGAGACCTAGATCCCTTA 176
|||||

RESULT 4
US-09-864-761-30344
; Sequence 30344, Application US/09864761

```

; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30344
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
; OTHER INFORMATION: EST_HUMAN HT: AW583777.1, EVALUUE 3.00e-51
; OTHER INFORMATION: NT HT: Z70177.1, EVALUUE 2.90e-01
; OTHER INFORMATION: SWISSPROT HIT: Q9Y0I1, EVALUUE 7.40e-01
; US-09-864-761-18355

Query Match 35.38; Score 106; DB 10; Length 467;
Best Local Similarity 100.0%; Pred. No. 7e-44;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 AGTATGCTCAGCCTAGGAGGAATTTGGTACACTTCATCAAGATCTACCTTTA 254
|||||
DB 1 AGTATGCTCAGCCTAGGAGGAATTTGGTACACTTCATCAAGATCTACCTTTA 60
|||||

QY 255 AACCAAGAAGAGAAAGACGAGCTATGACACCTAGATCCTTTA 300
|||||

Db 61 RACCAGAAGAGAAAGACGAGCTATGACACCTAGATCCTTTA 106
|||||

RESULT 5
US-09-864-761-18355
; Sequence 18355, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 18355
; LENGTH: 423
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010133.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
; US-09-864-761-18355
```

Query Match 6.3%; Score 19; DB 10; Length 423;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 AGAAGAGGAGGAGGAGCA 277
|||||
DB 379 AGAAGAGGAGGAGGAGCA 397

RESULT 6

US-09-764-877-3815/c
; Sequence 3815, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3815
; LENGTH: 7105
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3815

Query Match 6.3%; Score 19; DB 10; Length 7105;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 ATCAGCTGAGTCCAGAGA 162
|||||
DB 306 ATCAGCTGAGTCCAGAGA 2988

RESULT 7

US-09-825-414-1
; Sequence 1, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 30365
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (29734)
; OTHER INFORMATION: n at any position is undefined
US-09-825-414-1

Query Match 6.3%; Score 19; DB 10; Length 30365;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 AAAAGCAGTGGACCATGAG 109
|||||
DB 16404 AAAAGCAGTGGACCATGAG 16422

RESULT 8

US-09-867-701-490
; Sequence 490, Application US/09867701
; Patent No. US2002013237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 490
; LENGTH: 370
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-490

Query Match 6.0%; Score 18; DB 10; Length 370;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACAGAAAGAAAAAAC 18
|||||
DB 263 AAACAGAAAGAAAAAAC 280

RESULT 9

US-09-770-444-644
; Sequence 644, Application US/09770444
; Patent No. US20020023280A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2027 (PARA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 644
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:

```
; NAME/KEY: misc_feature
; LOCATION: (1)...(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-444-644

Query Match          6.0%; Score 18; DB 10; Length 452;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACAGAAAGAAAAAC 18
   |||||
Db 167 AAACAGAAAGAAAAAC 184

RESULT 10
US-09-864-761-14920
; Sequence 14920, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14920
; LENGTH: 581
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022327.17
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6

; NAME/KEY: misc_feature
; LOCATION: (1)...(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-444-644

Query Match          6.0%; Score 18; DB 10; Length 452;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAACAGAAAGAAAAAC 18
   |||||
Db 167 AAACAGAAAGAAAAAC 184

RESULT 11
US-09-938-842A-1349
; Sequence 1349, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAI
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1349
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1349

Query Match          6.0%; Score 18; DB 9; Length 735;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 AGAAGAGGAGAAAGAGC 276
   |||||
Db 195 AGAAGAGGAGAAAGAGC 212

RESULT 12
US-09-070-927A-506/c
; Sequence 506, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; APPLICANT: Patrick J. Dillon
; APPLICANT: Steven Barash
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypept:
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
```

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
SEQUENCE CHARACTERISTICS:
LENGTH: 789 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 506:
US-09-070-927A-506

Query Match 6.0%; Score 18; DB 10; Length 789;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 261 AAGAAGGAAAGAGCAG 278
|||||
Db 183 AAGAAGGAAAGAGCAG 166

RESULT 13

US-09-864-761-27614/C
; Sequence 27614, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27614
; LENGTH: 123
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005686.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.77
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.68
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.71
; OTHER INFORMATION: EST_HUMAN HIT: AI863650.1, EVALUE 6.30e+00
; OTHER INFORMATION: NT HIT: AF059679.1, EVALUE 4.30e+00
US-09-864-761-27614

Query Match 5.7%; Score 17; DB 10; Length 123;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 259 AGAAGAGGAAAGAG 275
|||||
Db 73 AGAAGAGGAAAGAG 57

RESULT 14

US-09-878-574-8457/C
; Sequence 8457, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Thompson, Michael J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules associated with
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 8457
; LENGTH: 171
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701101232H1
US-09-878-574-8457

Query Match 5.7%; Score 17; DB 10; Length 171;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACACAGAAAGAAAAA 17
|||||
Db 156 AACACAGAAAGAAAAA 140

RESULT 15

US-09-864-761-30968/C
; Sequence 30968, Application US/09864761
; Patent No. US20020048763A1

GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Acomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 30968
LENGTH: 224
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004104.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 52
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.67
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.54
OTHER INFORMATION: SWISSPROT HIT: P22544, EVALUE 4.40e-01
OTHER INFORMATION: EST_HUMAN HIT: AA189080.1, EVALUE 5.00e-33
US-09-864-761-30968

Query Match 5.7%; Score 17; DB 10; Length 224;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 258 CAGAGAGGAGAAAGAA 274
DB 31 CAGAGAGGAGAAAGAA 15
|||||

Search completed: January 21, 2003, 14:46:36
Job time : 47.6299 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 13:50:07 ; Search time 32.136 Seconds
(without alignments)
2488.651 Million cell updates/sec

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	99	100.0	669	21	AAV70539	Human Factor 8 Hom
2	76	76.8	197	22	ABG01298	Novel human diagno
3	19	19.2	583	22	AAU75450	Human colon cancer
4	8	8.1	100	22	AAU60183	Propionibacterium
C 5	8	8.2	242	22	ABB30277	Peptide #2928 enco
C 6	8	8.2	242	22	ABB35443	Peptide #2949 enco
C 7	8	8.2	242	22	ABB20882	Protein #2881 enco
C 8	8	8.2	242	22	AAU56266	Human brain expres
C 9	8	8.2	242	22	AAU68643	Human bone marrow
C 10	8	8.2	242	22	AAU16455	Peptide #2889 enco
C 11	8	8.2	242	22	AAU38950	Peptide #2987 enco
C 12	8	8.2	242	22	AAU04184	Peptide #2866 enco
C 13	8	8.2	242	23	ABG38222	Human peptide enco
14	8	8.1	260	22	ABB09212	Mouse GPI-anchored
15	8	8.1	260	22	AAU62106	Mouse RetL5 altern
16	8	8.1	277	22	AAU62103	Mouse RetL5 polype
17	8	8.1	280	22	ABB65868	Drosophila melanog
18	8	8.1	280	22	ABB71693	Drosophila melanog
19	8	8.1	293	22	ABB09215	Mouse putative tra
20	8	8.1	340	20	AAU42771	Murine GliR deriv
C 21	8	8.2	448	19	AAU82551	Human LIR-pbm2 pro
C 22	8	8.2	448	19	AAU53463	Human gp49 HM18 po
C 23	8	8.2	448	21	AAU04176	Leukocyte immunogl
C 24	8	8.2	472	19	AAU69234	PCR-IV protein seq
25	8	8.1	476	22	AAU62107	Murine RetL5/human
C 26	8	8.2	574	20	AAU14539	Rat lipolysis stim
C 27	8	8.2	574	22	AAU59905	Rat leptin fragmen
28	8	8.1	583	23	ABB92834	Herbicidally activ
C 29	8	8.2	593	20	AAU14538	Rat lipolysis stim
C 30	8	8.2	593	22	AAU59904	Rat leptin fragmen
31	7	7.1	9	18	AAU37009	TRP-2 derived pote
32	7	7.1	10	13	AAU38137	Cell-to-cell bindi
C 33	7	7.2	10	15	AAU38193	Hepatitis C virus
34	7	7.1	10	18	AAU30829	TRP-2 derived pote
C 35	7	7.2	10	20	AAU45761	Immunogenic peptid
36	7	7.1	11	13	AAU28090	Cell-to-cell bindi
C 37	7	7.2	14	23	ABP46389	Human BLYS binding
C 38	7	7.2	18	21	AAU51856	Human secreted pro
39	7	7.1	25	22	AAU61650	TRP2 peptide #2.
40	7	7.1	30	23	AAU84936	Human Trp2 segment
41	7	7.1	30	23	AAU84937	Human Trp2 segment
42	7	7.1	31	22	ABB43446	Peptide #10952 enco
43	7	7.1	31	22	ABB26414	Protein #8413 enco
44	7	7.1	31	22	AAU64375	Human brain expres
45	7	7.1	31	22	AAU77198	Human bone marrow

ALIGNMENTS

RESULT 1
AAAY70539
ID AAY70539 standard: Protein: 669 AA:

AA
AC
AA70539;

DT 04-JUL-2000 (first entry)

DE Human Factor 8 Homologue.

KW	Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW	cerebroprotective; therapeutic; coagulation related disorder;
KW	haemophilia; stroke; screening.

OS Homo sapiens.

PN WO200012532-A1.

PD 09-MAR-2000.

QY 251 TTTAAACCAAGAGAGGAAACAGCAGCTATGACCTAGACCTAGATCCT 298
|||||
Db 61 PheLysProGluGlyLysGluAlaGlyTyrAlaAspLeuAspPro 76

RESULT 3
AAG75450
ID AAG75450 standard; Protein; 583 AA.

XX AAG75450;
XX
XX 03-SEP-2001 (first entry)
XX Human colon cancer antigen protein SEQ ID NO:6214.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX Homo sapiens.

OS
XX WO200122920-A2.
XX
XX PD 05-APR-2001.
XX
XX PF 28-SEP-2000; 2000WO-US26524.
XX
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/24.
XX N-PSDB; AAH34855.
XX

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 7657-7660; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAB7789 represent sequences used in the exemplification of the
XX present invention.

XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 583 AA;

Alignment Scores:
Pred. No.: 1,01e-09 Length: 583
Score: 19,00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19,19% Indels: 0
DB: 22 Gaps: 0

US-10-060-830-4 (1-300) x AAG75450 (1-583)

QY 2 AACACAAAGAAACACTGAAGCAGCTATGACCTAGACCTACTGGACCGGCAGGT 58
|||||
Db 534 AsnArgLysLysLysThrGluGlyThrTyraSpLeuProTyrTrpAspArgAlaGly 552

RESULT 4
AAU60183
ID AAU60183 standard; Protein; 100 AA.

XX AAU60183;
XX
XX 27-FEB-2002 (first entry)
XX Propionibacterium acnes immunogenic protein #21079.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.
XX
XX PD 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIYA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.
XX N-PSDB; AAS59608.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris -

XX Example 1; SEQ ID NO 21378; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).

XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO.
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 100 AA;

Alignment Scores:
Pred. No.: 39,2 Length: 100
Score: 8,00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 8.08% Indels: 0
DB: 22 Gaps: 0
US-10-060-830-4 (1-300) x AAU60183 (1-100)
QY . 196 GTATGCTCAGCCACTGCTAGGAGG 219
DB 69 ValCysSerAlaThrGlyArg 76
RESULT 5
ABB30277
ID ABB30277 standard; Peptide; 242 AA.
XX
AC ABB30277;
XX
DT 01-FEB-2002 (first entry)
XX
DE Peptide #2928 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes
XX
PS Claim 27; SEQ ID NO 13245; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and Bt 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a peptide encoded by a single exon nucleic acid
XX probe of the invention.
CC Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 242 AA;
Alignment Scores:
Pred. No.: 33.2 Length: 242
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 22 Gaps: 0

Pred. No.: 33.2 Length: 242
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 22 Gaps: 0
US-10-060-830-4 (1-300) x ABB30277 (1-242)
QY 281 AGCTGCTCTCTTTCTCTCTCTG 258
DB 52 SerLeuLeuPheLeuLeuLeu 59
RESULT 6
ABB35443
ID ABB35443 standard; Peptide; 242 AA.
XX
AC ABB35443;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #2949 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver
XX
PS Claim 27; SEQ ID NO 28078; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 242 AA;
Alignment Scores:
Pred. No.: 33.2 Length: 242
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 22 Gaps: 0
US-10-060-830-4 (1-300) x ABB35443 (1-242)

```

Qy 281 AGCGTCTTCTTTCTCTTCTG 258
Db 52 SerLeuLeuPheLeuLeu 59
RESULT 7
ABB20882
ID ABB20882 standard; Protein; 242 AA.
AC ABB20882;
XX
XX
XX 23-JAN-2002 (first entry)
XX
XX Protein #2881 encoded by probe for measuring heart cell gene expression.
DE Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
XX Homo sapiens.
OS
XX WO200157274-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00666.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488899/53.
DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
XX Claim 15; SEQ ID NO 22652; 530pp; English.
PS
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC AB21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 242 AA;
SQ
Alignment Scores:
Pred. No.: 33.2 Length: 242
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 22 Gaps: 0
US-10-060-830-4 (1-300) x ABB20882 (1-242)
Qy 281 AGCGTCTTCTTTCTCTTCTG 258
Db 52 SerLeuLeuPheLeuLeu 59
RESULT 9
AAM68643
ID AAM68643 standard; Protein; 242 AA.
XX
XX
XX AAM68643;
AC
XX
XX

```

```

Db 52 SerLeuLeuPheLeuLeu 59
RESULT 8
AAM56266
ID AAM56266 standard; Protein; 242 AA.
XX
XX AAM56266;
AC
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 28371.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
XX Homo sapiens.
OS
XX WO200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483446/52.
DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
XX Example 4; SEQ ID NO: 28371; 650pp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, of
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
XX Sequence 242 AA;
SQ
Alignment Scores:
Pred. No.: 33.2 Length: 242
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 22 Gaps: 0
US-10-060-830-4 (1-300) x AAM56266 (1-242)
Qy 281 AGCGTCTTCTTTCTCTTCTG 258
Db 52 SerLeuLeuPheLeuLeu 59
RESULT 9
AAM68643
ID AAM68643 standard; Protein; 242 AA.
XX
XX
XX AAM68643;
AC
XX
XX

```

```

-DT XX 06-NOV-2001 (first entry)
DE XX Human bone marrow expressed probe encoded protein SEQ ID NO: 28949.
XX XX
KW KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX XX
OS OS Homo sapiens.
XX XX
PN PN WO200157276-A2.
XX XX
PD PD 09-AUG-2001.
XX XX
XX 30-JAN-2001; 2001WO-US00668.
XX XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX XX
PS Example 4; SEQ ID NO: 28949; 658pp + Sequence Listing; English.
XX XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX XX
SQ Sequence 242 AA;

Alignment Scores:
Pred. No.: 33.2 Length: 242
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 22 Gaps: 0

US-10-060-830-4 (1-300) x AAM68643 (1-242)
QY 281 AGCTGCTCTTCTTCTTCTTCTG 258
DB 52 SerLeuLeuPheLeuLeuLeu 59

RESULT 10
AAM16455
ID AAM16455 standard; Protein; 242 AA.
XX AC AAM16455;
XX XX
DT 12-OCT-2001 (first entry)
XX XX
DE Peptide #2889 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.
XX OS Homo sapiens.
XX PN WO200157278-A2.

```

```

XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00670.
XX XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX XX
PS Claim 27; SEQ ID NO 21281; 487pp; English.
XX XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 242 AA;

Alignment Scores:
Pred. No.: 33.2 Length: 242
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 22 Gaps: 0

US-10-060-830-4 (1-300) x AAM16455 (1-242)
QY 281 AGCTGCTCTTCTTCTTCTTCTG 258
DB 52 SerLeuLeuPheLeuLeuLeu 59

RESULT 11
AAM28950
ID AAM28950 standard; Protein; 242 AA.
XX AC AAM28950;
XX XX
DT 17-OCT-2001 (first entry)
XX XX
DE Peptide #2987 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX XX
PD 09-AUG-2001.
XX XX
XX 30-JAN-2001; 2001WO-US00663.

```

PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 27; SEQ ID NO 29219; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP;
 CC see AA131315-AA157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 SQ Sequence 242 AA;
 Alignment Scores:
 Pred. No.: 33.2 Length: 242
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 22 Gaps: 0
 US-10-060-830-4 (1-300) x AAM28950 (1-242)
 QY 281 AGCGTCTCTTTCTCTCTCTG 258
 Db 52 SerLeuLeuPheLeuLeuLeu 59
 RESULT 12
 AAM04184
 ID AAM04184 standard; Protein; 242 AA.
 XX
 AC AAM04184;
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Peptide #2866 encoded by probe for measuring breast gene expression.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 XX Homo sapiens.
 OS
 XX WO200157270-A2.
 PN
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US00661.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-476286/51.
 DR
 XX Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 27; SEQ ID NO 12924; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes
 CC (see AA100010-AA110067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of development,
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 242 AA;
 Alignment Scores:
 Pred. No.: 33.2 Length: 242
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 22 Gaps: 0
 US-10-060-830-4 (1-300) x AAM04184 (1-242)
 QY 281 AGCGTCTCTTTCTCTCTCTG 258
 Db 52 SerLeuLeuPheLeuLeuLeu 59
 RESULT 13
 ABG38222
 ID ABG38222 standard; Peptide; 242 AA.
 XX
 AC ABG38222;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 27887.
 XX
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 XX 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-234687P.
 PR 04-OCT-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2002-114183/15.
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX Claim 27; SEQ ID NO 27887; 634pp; English.
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 242 AA;
 Alignment Scores:
 Pred. No.: 33.2 Length: 242
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 23 Gaps: 0

US-10-060-830-4 (1-300) x ABG38222 (1-242)

OY 281 AGCGTCTCTTTCTCTCTCTG 258
 |||||

Db 52 SerLeuLeuPheLeuLeu 59

RESULT 14
 ABB09214

ID ABB09214 standard; Protein; 260 AA.

XX SQ Sequence 260 AA;
 Alignment Scores:
 Pred. No.: 32.8 Length: 260
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.08% Indels: 0
 DB: 22 Gaps: 0

US-10-060-830-4 (1-300) x ABB09214 (1-260)

OY 271 AGAAGCAGCGCTATGCAGACCTAGA 294
 |||||

Db 132 ArgSerArgLeuCysArgProArg 139

XX AC ABB09214;
 XX 08-JUL-2002 (first entry)
 XX Mouse GPI-anchored isoform a1 protein SEQ ID NO:1.
 XX GFRalpha4; glycosyl-phosphatidylinositol; GPI; GDNF; cytostatic;
 KW glycosyl-phosphatidylinositol-linked GDNF family alpha-receptor;
 KW glial cell line derived neurotrophic factor; osteopathic; tumour;
 KW neuroprotective; anticonvulsant; neoplasia; endocrine tumour;
 KW medullary thyroid carcinoma; pheochromocytoma; parathyroid hyperplasia;
 KW neuronal disorder; aberrant axonal sprouting.
 XX Mus musculus.
 OS WO200162795-A1.
 XX 30-AUG-2001.
 XX 14-NOV-2000; 2000WO-FI00994.
 XX 21-FEB-2000; 2000FI-0000394.
 XX (LICE-) LICENTIA LTD.
 XX Airaksinen M, Saarna M, Poteriaev D, Lindahl M, Timmusk T;
 PI Rossi J;
 XX WPI; 2001-596722/67.
 DR N-PSDB; ABL51669.
 XX New nucleic acid sequence for manufacturing polypeptides for treating
 PT endocrine cancers comprises a cDNA encoding a splicing isoform of
 PT mammalian growth factor receptor (GFR)alpha4 -
 XX Claim 9; Fig 18B; 143pp; English.
 XX The present invention describes an isolated and purified cDNA sequence
 CC encoding a splicing isoform of a mammalian growth factor receptor
 CC (GFR)alpha4, or its fragments. GFRalpha4 sequences have cytostatic,
 CC osteoprotective, neuroprotective and anticonvulsant activities. GFRalpha4 is
 CC a glycosyl-phosphatidylinositol (GPI)-linked glial cell line-derived
 CC neurotrophic factor (GDNF) family alpha-receptor. A GFRalpha4
 CC polynucleotide sequence can be used for recording GFRalpha4 mediated
 CC signalling in neurons or endocrine cells such as thyroid calcitonin-
 CC producing C-cells, parathyroid gland cells, adrenal chromaffin cells, or
 CC cells from the pituitary intermediate lobe. GFRalpha4 protein and
 CC polynucleotide sequences can be used for manufacturing polypeptides
 CC useful for diagnosing and/or treating tumours in parathyroid gland cells,
 CC adrenal chromaffin cells, cells of pituitary intermediate lobe,
 CC neoplasia, endocrine tumours, medullary thyroid carcinoma and
 CC pheochromocytoma, parathyroid hyperplasia, neuronal disorders or for
 CC preventing neuronal death or aberrant axonal sprouting. The present
 CC sequence represents the mouse GFRalpha4 protein, designated GPI-anchored
 CC isoform a1, from the present invention.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:49:12 ; Search time 16.063 Seconds
(without alignments)
3590.903 Million cell updates/sec

Title: US-10-060-830-4

Perfect score: 99

Sequence: 1 aaacagaagaaaaaactg.....tatgcagacctagatcctta 300

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565536

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US10060830/runat_16012003.092703.1513/app_query.fasta.1.1109
-DB=PIR_73 -QFMT=fastan -SUFFIX=olin2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NOCALIGN=200 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn.1.1.26.erunat_16012003.092703.1513 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	9	9.1	3147	T21328	hypothetical prote
2	8	8.1	104	H90197	hypothetical prote
3	8	8.2	125	S62925	probable membrane
4	8	8.2	209	B42687	neurotrophin-4 pre
5	8	8.2	259	T34781	probable signal pe
6	8	8.2	342	T35790	probable transmemb
7	8	8.2	445	E22845	hypothetical prote
8	8	8.2	529	T24231	hypothetical prote
9	8	8.1	583	T01470	diphosphate-fructo
10	8	8.1	649	AE3265	hypothetical prote
11	8	8.1	1275	AD0332	probable membrane
12	7	7.2	44	1 W5WLEB	E5 protein - bovin
13	7	7.2	44	1 W5WLB2	E5 protein - bovin
14	7	7.2	51	2 S64676	acetylcholinestera

15	7	7.1	93	2	T17982	hypothetical prote
16	7	7.2	103	2	E72711	hypothetical prote
17	7	7.2	104	2	B81787	probable inner mem
18	7	7.2	133	2	B30902	probable membrane
19	7	7.2	153	2	AB1633	hypothetical prote
20	7	7.2	153	2	AH1270	hypothetical prote
21	7	7.2	160	2	E81211	hypothetical prote
22	7	7.2	174	2	G70220	exported protein A
23	7	7.2	185	2	T01914	probable disease r
24	7	7.2	186	2	F82945	hypothetical prote
25	7	7.2	189	2	H70416	hypothetical prote
26	7	7.2	193	2	T34047	hypothetical prote
27	7	7.2	196	2	T27833	hypothetical prote
28	7	7.1	196	2	S50487	hypothetical prote
29	7	7.2	197	2	T47848	hypothetical prote
30	7	7.2	205	2	T10384	hypothetical prote
31	7	7.1	206	2	B48441	antigen (C-termina
32	7	7.2	214	2	C89837	conserved hypothet
33	7	7.2	219	2	A35650	Sur protein - chic
34	7	7.2	242	2	AG1868	hypothetical prote
35	7	7.2	246	2	B72728	probable ribosomal
36	7	7.2	250	2	T36131	hypothetical prote
37	7	7.1	257	1	C64812	molybdate-binding
38	7	7.1	257	2	G90727	molybdate-binding
39	7	7.1	257	2	H85578	molybdate-binding
40	7	7.1	257	2	AC0595	molybdate-binding
41	7	7.2	260	2	D86266	hypothetical prote
42	7	7.2	266	2	T05471	hypothetical prote
43	7	7.1	273	2	A32123	H+-exporting ATPas
44	7	7.1	274	2	JN0508	H+-exporting ATPas
45	7	7.2	281	2	G71095	hypothetical prote

ALIGNMENTS

RESULT 1

T21328

hypothetical protein F25C8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C:Accession: T21328

R:Matthews, L.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19407

A:Accession: T21328

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-3147 <WIL>

A:Cross-references: EMBL:281512; PIDN:CAB04172.1; GSPDB:GN00023; CESP:F25C8.3

A:Experimental source: clone F25C8

C:Genetics:

A:Gene: CESP:F25C8.3

A:Map position: 5

A:Introns: 32/2; 57/2; 73/3; 113/3; 157/3; 366/1; 456/1; 509/1; 638/1; 773/2; 848/3; 2510/2; 2737/3; 2827/1; 2902/3; 3093/3; 3121/1

C:Superfamily: Caenorhabditis elegans hypothetical protein F25C8.3

Alignment Scores:

Pred. No.:	0.542	Length:	3147
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	9.09%	Indels:	0
DB:	2	Gaps:	0

US-10-060-830-4 (1-300) x T21328 (1-3147)

QY 145 TCACCTGAGTCCAGAGAGTCCACCAC 171

|||||

DB 2403 SerProGluSerLysArgSerHisHis 2411

RESULT 2

H90197

hypothetical protein SSO0521 [imported] - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
 C:Accession: H90197
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awaye, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.
 arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001
 A:Description: Sulfolobus solfataricus complete genome.
 A:Reference number: A99139
 A:Accession: H90197
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-104 <KUR>
 A:Cross-references: GB:AE006641; NID:gi13813679; PIDN:AAK40839.1; GSPDB:GN00155
 C:Genetics:
 A:Gene: SSO0521

Alignment Scores:
 Pred. No.: 9.13 Length: 104
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.08% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-4 (1-300) x H90197 (1-104)
 QY 206 CCACGTGGTAGGAGGATTCGTG 229
 Db 86 ProLeuValGlyGlyLeuValGly 93
 |||||||||||||||||||

RESULT 3
 S62925
 probable membrane protein YNL013c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein N2854
 C:Species: Saccharomyces cerevisiae
 C>Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
 C:Accession: S62925
 R:Andre, B.; Iraqui Houssaini, I.; Urrestarazu, L.A.; Vissers, S.
 submitted to the Protein Sequence Database, April 1996
 A:Reference number: S62920
 A:Accession: S62925
 A:Molecule type: DNA
 A:Residues: 1-125 <AND>
 A:Cross-references: EMBL:Z71290; NID:gi1301835; PID:gi1301837; GSPDB:GN00014; MIPS:YNL013C
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: MIPS:YNL013c
 A:Cross-references: SGD:S0004958
 A:Map position: 14L
 C:Superfamily: Saccharomyces probable membrane protein YNL013c
 C:Keywords: transmembrane protein
 F:28-44/Domain: transmembrane #status predicted <TM>

Alignment Scores:
 Pred. No.: 8.94 Length: 125
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-4 (1-300) x S62925 (1-125)
 QY 287 CTGCATAGCTGCTCTTCTTCTT 264
 Db 37 leuHisSerLeuLeuLeuPheLeu 44
 |||||||||||||||||||

RESULT 4
 B42687
 neurotrophin-4 precursor - rat
 C:Species: Rattus norvegicus (Norway rat)

C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999
 C:Accession: B42687; JH0504; JH0505
 R:IP, N.Y.; Ibanez, C.F.; Nye, S.H.; McClain, J.; Jones, P.F.; Gies, D.R.; Belluscio
 Proc. Natl. Acad. Sci. U.S.A. 89, 3060-3064, 1992
 A:Title: Mammalian neurotrophin-4: structure, chromosomal localization, tissue distr
 A:Reference number: A42687; MUID:92212967; PMID:1313578
 A:Accession: B42687
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-209 <IPA>
 A:Cross-references: GB:M86742; NID:g205775; PIDN:AAA41728.1; PID:g205776
 R:Berkemeier, L.R.; Winslow, J.W.; Kaplan, D.R.; Nikolics, K.; Goeddel, D.V.; Rosent
 Neuron 7, 857-866, 1991
 A:Title: Neurotrophin-5: a novel neurotrophic factor that activates trk and trkB.
 A:Reference number: JH0503; MUID:92075279; PMID:1742028
 A:Accession: JH0504
 A:Molecule type: DNA
 A:Residues: 1-209 <BER>
 A:Accession: JH0505
 A:Molecule type: mRNA
 A:Residues: 1-176, 'p', 178-209 <BER1>
 A:Cross-references: GB:S69323; NID:g240025; PIDN:AAB20548.1; PID:g240026
 C:Comment: This protein is a targeted-derived, diffusible neurotrophic factor.
 C:Comment: The neurotrophins stimulate autophosphorylation and transduce signals thr
 C:Superfamily: nerve growth factor beta chain
 C:Keywords: glycoprotein
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-79/Domain: propeptide #status predicted <PRO>
 F:80-209/Product: neurotrophin-5 #status predicted <NEU>
 F:75/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
 Pred. No.: 8.43 Length: 209
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-4 (1-300) x B42687 (1-209)
 QY 281 AGCTGCTCTTCTTCTTCTTCTG 258
 Db 8 SerLeuLeuPheLeuLeuLeu 15
 |||||||||||||||||||

RESULT 5
 T34781
 probable signal peptidase I (EC 3.4.21.89) sipl SC2E1.13 [similarity] - Streptomyces
 C:Species: Streptomyces coelicolor
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 24-May-2001
 C:Accession: T34781
 R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1998
 A:Reference number: Z21557
 A:Accession: T34781
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-259 <MUR>
 A:Cross-references: EMBL:AL023797; PIDN:CAAL9388.1; GSPDB:GN00070; SCOEDB:SC2E1.13
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: sipl; SCOEDB:SC2E1.13
 C:Superfamily: signal peptidase I
 C:Keywords: hydrolase; serine proteinase

Alignment Scores:
 Pred. No.: 8.22 Length: 259
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 2 Gaps: 0

US-10-060-830-4 (1-300) x T34781 (1-259)

QY 278 CTGCTCTCTTTCTCTCTCTCTGTT 255
|||||
Db 56 LeuLeuLeuPheLeuLeuVal 63

RESULT 6

T35790

probable transmembrane sugar transport protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 18-Feb-2000

C:Accession: T35790

R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1998

A:Reference number: 221570

A:Accession: T35790

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-342 <SEE>

A:Cross-references: EMBL:AL031013; PIDN:CAA19795.1; GSPDB:GN00070; SCOEDB:SC8A6.23

A:Experimental source: strain A3(2)

C:Genetics:

C:Superfamily: inner membrane protein upgA

Alignment Scores:
Pred. No.: 7.96 Length: 342
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-4 (1-300) x T35790 (1-342)

QY 278 CTGCTCTCTTTCTCTCTCTCTGTT 255
|||||
Db 309 LeuLeuLeuPheLeuLeuVal 316

RESULT 7

E22845

hypothetical protein 4 - Trypanosoma brucei mitochondrion

C:Species: mitochondrion Trypanosoma brucei

C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 11-May-2000

C:Accession: E22845

R:Hensgens, L.A.M.; Brakenhoff, J.; De Vries, B.F.; Sloof, P.; Tromp, M.C.; Van Boom, J.

Nucleic Acids Res. 12, 7327-7344, 1984

A:Title: The sequence of the gene for cytochrome c oxidase subunit I, a frameshift conta

ondial maxi-circle DNA

A:Reference number: A93537; MUID:85037915; PMID:6093040

A:Accession: E22845

A:Molecule type: DNA

A:Residues: 1-445 <HEN>

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SGC6

C:Superfamily: hypothetical protein 1 (Cyb-COII intergenic region)

C:Keywords: mitochondrion

Alignment Scores:
Pred. No.: 7.73 Length: 445
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-4 (1-300) x E22845 (1-445)

QY 281 AGCCTGCTCTCTTTCTCTCTCTG 258
|||||
Db 112 SerLeuLeuPheLeuLeu 119

RESULT 8

T24231

hypothetical protein R166.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T24231

R:Matthews, P.

submitted to the EMBL Data Library, August 1995

A:Reference number: Z19859

A:Accession: T24231

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-529 <WIL>

A:Cross-references: EMBL:Z50795; PIDN:CAA90666.1; GSPDB:GN00020; CESP:R166.4

A:Experimental source: clone R166

C:Genetics:

A:Gene: CESP:R166.4

A:Map position: 2

A:Introns: 18/3; 63/1; 325/3

Alignment Scores:
Pred. No.: 7.58 Length: 529
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-4 (1-300) x T24231 (1-529)

QY 123 CGAACTGGGGTTTCCTCATGGTCC 100
|||||
Db 38 ArgThrGlyValSerSerIrrpSer 45

RESULT 9

T01470

diphosphate-fructose-6-phosphate 1-phosphotransferase (EC 2.7.1.90) beta chain -

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 03-Jun-2002

C:Accession: T01470

R:Courtney, L.; Stoneking, T.; Langston, Y.; Mead, K.

submitted to the EMBL Data Library, August 1998

A:Description: The sequence of A. thaliana T24H24.

A:Reference number: Z14333

A:Accession: T01470

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-583 <COU>

A:Cross-references: EMBL:AF075598; NID:G3293581; PIDN:AAC28214.1; PID:G3377841

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 67/1; 121/1; 177/3; 195/3; 231/2; 268/1; 286/1; 315/3; 359/3; 416/3; 4

A>Note: T24H24.15

C:Superfamily: pyrophosphate-fructose-6-phosphate 1-phosphotransferase alpha cha:

C:Keywords: Phosphotransferase

Alignment Scores:
Pred. No.: 7.49 Length: 583
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.08% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-4 (1-300) x T01470 (1-583)

QY 8 AAGAAAAAACTGAAGCACCTAT 31
|||||
Db 436 LysLysLysThrGluGlyThrTyr 443

RESULT 10

AE2265

hypothetical protein alr3676 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AE2265
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena Res. 8, 205-213, 2001
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2265
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-649 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075375.1; PID:g17132809; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3676

Alignment Scores:
Pred. No.: 7.4 Length: 649
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.08% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-4 (1-300) x AE2265 (1-649)

QY 123 GCTATAGCAGCAGCGAAGTAAATC 146
|||||
Db 518 AlAtleAlaAlaAlaLysLeulle 525

RESULT 11

AD0332

Probable membrane protein YPO2724 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AD0332
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Ill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0332
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1275 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC29263.1; PID:g15980702; GSPDB:GN00175
C:Genetics:
A:Gene: YPO2724

Alignment Scores:
Pred. No.: 6.85 Length: 1275
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.08% Indels: 0
DB: 2 Gaps: 0

US-10-060-830-4 (1-300) x AD0332 (1-1275)

QY 182 GCTGACTGCGAGAGTATGCTCAG 205
|||||
Db 360 AlaAspSerAlaGluTyAlaGln 867

RESULT 12

W5WLB2

E5 protein - bovine papillomavirus type 1

C:Species: bovine papillomavirus type 1
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 28-Jul-2000
C:Accession: G31169; E18151; S12366; B18151

R:Chen, E.Y.; Howley, P.M.; Levinson, A.D.; Seeburg, P.H.

Nature 299, 529-534, 1982
A:Title: The primary structure and genetic organization of the bovine papillomavirus
A:Reference number: A93289; MUID:83012974; PMID:6289124
A:Accession: G31169
A:Molecule type: DNA
A:Residues: 1-44 <CH>
A:Cross-references: GB:X02346; GB:J02044; GB:M24622; GB:X00473; NID:g60965; PIDN:CA
R:Schlegel, R.; Wade-Glass, M.; Rabson, M.S.; Yang, Y.C. Science 233, 464-467, 1986
A:Title: The E5 transforming gene of bovine papillomavirus encodes a small, hydrophobic
A:Reference number: A94289; MUID:86261794; PMID:3014660
A:Accession: E18151
A:Molecule type: protein
A:Residues: 1-44 <SCH>
R:Goldstein, D.J.; Schlegel, R. EMBO J. 9, 137-146, 1990
A:Title: The E5 oncoprotein of bovine papillomavirus binds to a 16 kd cellular protein
A:Reference number: S12366; MUID:90107936; PMID:1688529
A:Accession: S12366
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-44 <GOL>
C:Comment: This protein is a small, membrane-associated, hydrophobic polypeptide.
C:Superfamily: bovine papillomavirus E5 protein
C:Keywords: early protein; membrane protein; transforming protein

Alignment Scores:
Pred. No.: 115 Length: 44
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x W5WLB2 (1-44)

QY 278 CTGCTCTTTTCCCTCTCTG 258
|||||
Db 20 LeuLeuLeuPheLeuLeuLeu 26

RESULT 13

W5WLB2

E5 protein - bovine papillomavirus type 2
C:Species: bovine papillomavirus type 2
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C:Accession: F18151; B18151; G31169
R:Groff, D.E.; Mitra, R.; Lancaster, W.D. submitted to GenBank, May 1988
A:Reference number: A94519
A:Accession: F18151
A:Molecule type: DNA
A:Residues: 1-44 <GRO>
A:Cross-references: GB:M20219; GB:M19551; NID:g332996; PIDN:AAA66838.1; PID:g808785
C:Comment: This protein is a small, membrane-associated, hydrophobic polypeptide.
C:Superfamily: bovine papillomavirus E5 protein
C:Keywords: early protein; membrane protein; transforming protein

Alignment Scores:
Pred. No.: 115 Length: 44
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x W5WLB2 (1-44)

QY 278 CTGCTCTTTTCCCTCTCTG 258
|||||
Db 20 LeuLeuLeuPheLeuLeuLeu 26

RESULT 14

S64676
 acetylcholinesterase (EC 3.1.1.7) - bovine (fragment)
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 07-May-1999
 C:Accession: S64676; S64677
 R:Haas, R.; Jackson, B.C.; Reinhold, B.; Foster, J.D.; Rosenberry, T.L.
 Biochem. J. 314, 817-825, 1996
 A:Title: Glycylserine phospholipid anchor and protein C-terminus of bovine erythrocyte
 A:Reference number: S64676; MUID:96177853; PMID:8615775
 A:Accession: S64676
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-51 <HAA>
 A:Accession: S64677
 A:Molecule type: protein
 A:Residues: 1-23 <HAW>
 C:Keywords: carboxylic ester hydrolase

Alignment Scores:			
Pred. No.:	113	Length:	51
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	7.22%	Indels:	0
DB:	2	Gaps:	0

US-10-060-830-4 (1-300) x S64676 (1-51)

QY 278 CTGCTCTTTCTCTCTCTG 258

Db 39 LeuLeuLeuPheLeuLeuLeu 45

RESULT 15

TI7982

hypothetical protein A480L - Chlorella virus PBCV-1

C:Species: Chlorella virus PBCV-1

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: TI7982

R:Graves, M.V.; Van Etten, J.L.

submitted to the EMBL Data Library, May 1999

A:Reference number: Z18806

A:Accession: TI7982

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-93 <GRA>

A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96847.1

A:Experimental source: specific host Chlorella strain NC64A

C:Genetics:

A:Note: A480L

Alignment Scores:

Pred. No.:	105	Length:	93
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	7.07%	Indels:	0
DB:	2	Gaps:	0

US-10-060-830-4 (1-300) x TI7982 (1-93)

QY 34 CTTACCTTACTGGACCGGC 54

Db 67 LeuThrLeuLeuGlyProGly 73

Search completed: January 21, 2003, 14:54:43
 Job time : 20.063 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 13:52:06 ; Search time 9.2126 Seconds
(without alignments)
2701.279 Million cell updates/sec

Title: US-10-060-830-4

Perfect score: 99

Sequence: 1 aaacagagaagaaactg.....tatgcagacctagatcctta 300

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Word size: 1

Total number of hits satisfying chosen parameters: 225614

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -MODEL-frame+n2p model -DEV-xmlh
-Q/cgn2_1/USPTO.spool/US10060830/runat_16012003_092701_1487/app_query.fasta_1.1109
-DB-SwissProt_40 -Qfmt-fastan -SUFFIX-olin2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-oligo -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn_1.1.1 @runat_16012003_092701_1487 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	8.2	125	1 YNB3_YEAST	P53979 saccharomyc
2	8	8.2	209	1 NT5_RAT	P34131 rattus norv
3	8	8.2	398	1 CYB_WHEAT	P07747 triticum ae
4	7	7.2	44	1 VES_BPV1	P06928 bovine pap
5	7	7.2	121	1 YJF1_YEAST	P14359 saccharomyc
6	7	7.2	186	1 Y007_UREPA	P09943 escherichia
7	7	7.2	189	1 YD48_AQUAE	P06736 aquifex aeo
8	7	7.1	196	1 YEL9_YEAST	P40018 saccharomyc
9	7	7.2	205	1 Y115_NPVOP	O10354 oryza pseu
10	7	7.2	246	1 RS3_AERPE	P09178 aeropyrum p
11	7	7.1	257	1 MODA_ECOLI	P37329 escherichia
12	7	7.2	283	1 PLCA_HUMAN	P09943 homo sapien
13	7	7.1	311	1 PRMA_CLOAB	P45558 clostridium
14	7	7.1	351	1 VAOD_HUMAN	P12953 homo sapien
15	7	7.1	351	1 VAOD_MOUSE	P51863 mus musculu
16	7	7.1	359	1 GP15_MACFA	O9bd57 macaca fasc
17	7	7.1	360	1 GP15_CERAE	O18982 cercopithe
18	7	7.1	360	1 GP15_MACMU	O97663 macaca mula

19	7	7.1	360	1 GP15_MACNE	P56412 macaca neme
20	7	7.2	366	1 IHA_HUMAN	P05111 homo sapien
21	7	7.1	372	1 YD80_BACHD	O9kd39 bacillus ha
22	7	7.2	389	1 THIL_MYCTU	Q10629 mycobacteri
23	7	7.2	393	1 THIL_MYCLE	P46707 mycobacteri
24	7	7.1	447	1 MCRY_METJA	Q58252 methanococc
25	7	7.2	484	1 CATA_CANTR	P07820 candida tro
26	7	7.2	486	1 CATA_CANAL	O13289 candida alb
27	7	7.1	491	1 GALT_STRMU	P96994 streptococc
28	7	7.1	493	1 GALT_LACLA	Q9ce63 lactococcus
29	7	7.1	493	1 GALT_LACLC	Q9ce63 lactococcus
30	7	7.2	515	1 CATA_YEAST	P15202 saccharomyc
31	7	7.2	517	1 TYR2_MOUSE	P29812 mus musculu
32	7	7.1	519	1 TYR2_HUMAN	O40126 homo sapien
33	7	7.2	572	1 RECN_STRCO	Q9s220 streptomyce
34	7	7.1	616	1 MS2_ARATH	Q08891 arabidopsis
35	7	7.1	618	1 DLL3_HUMAN	Q9ny17 homo sapien
36	7	7.1	624	1 YH19_YEAST	P38900 saccharomyc
37	7	7.1	672	1 YCDR_ECOLI	P75906 escherichia
38	7	7.1	697	1 YHF0_YEAST	P38721 saccharomyc
39	7	7.2	731	1 GLGB_MYCTU	O10625 mycobacteri
40	7	7.1	731	1 SUF1_XENLA	Q9pnu2 xenopus lae
41	7	7.2	829	1 CAD3_HUMAN	P22223 homo sapien
42	7	7.2	854	1 PI4K_HUMAN	P42356 homo sapien
43	7	7.1	949	1 AHM6_ARATH	Q9sacz9 arabidopsis
44	7	7.2	993	1 VIA_CMVFN	P17769 cucurbit mo
45	7	7.2	993	1 VIA_CMVII	Q83270 cucurbit mo

ALIGNMENTS

RESULT 1
YNB3_YEAST
ID YNB3_YEAST STANDARD; PRT; 125 AA.
AC P53979;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 15.0 kDa protein in YEF3B-SPO1 intergenic region.
GN YNL013C OR N2854.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Andre B., Iraqi Housaini I., Urrestarazu L.A., Visser S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 271290; CAA95875.1; --
DR SGD; S0004958; YNL013C.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 28 48 POTENTIAL.
FT TRANSMEM 54 74 POTENTIAL.
SQ SEQUENCE 125 AA; 14973 MW; 8E7C3E847F445CEFCRC64;
Alignment Scores:
Pred. No.: 6.68 Length: 125
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 1 Gaps: 0
US-10-060-830-4 (1-300) x YNB3_YEAST (1-125)

Alignment Scores: 6.25 Length: 209
Pred. No.: 8.00 Matches: 8
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 8.25% Gaps: 0
DB:

US-10-060-830-4 (1-300) x NT5_RAT (1-209)

QY 281 AGCCTGCTCTCTTTTCCCTTCCTG 258
|||||

Db 8 SerLeuLeuPheLeuLeu 15

RESULT 3

CYB_WHEAT
ID CYB_WHEAT STANDARD; PRT: 398 AA.
AC P07747:
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cytochrome B.
GN COB OR CVTB.
OS Triticum aestivum (Wheat).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85215614; PubMed=2987849;
RA Boer P.H., McIntosh J.E., Gray M.W., Bonen L.;
RT "The wheat mitochondrial gene for apocytochrome b: absence of a
RT prokaryotic ribosome binding site."
RL Nucleic Acids Res. 13:2281-2292(1985).
RN [2]

RNA EDITING.
RX MEDLINE=90015192; PubMed=2552325;
RA Gualberto J.M., Lamattina L., Bonnard G., Weil J.-H.,
RA Grienenberger J.-M.;
RT "RNA editing in wheat mitochondria results in the conservation of
RT protein sequences".
RL Nature 341:660-662(1989).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -!- CAUTION: POSITIONS 96, 100, 109, 120, 227 AND 242 ARE MODIFIED DUE
CC TO RNA EDITING

This SWISS-PROT entry is copyright. It is produced through a collabora-
tion between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on
the use by non-profit institutions as long as its content is in no
way modified and this statement is not removed. Usage by and for commer-
cial entities requires a license agreement (See <http://www.isb-sib.ch/announcements>
or send an email to license@isb-sib.ch).

DB EMBL; X02352; CAA26207.1; ALT_SEQ.
DR PIR; A22931; A22931.
DR InterPro; IPR000179; Cyt_b.b6.
DR Pfam; PF00032; cytochrome_b_c1.
DR Pfam; PF00033; cytochrome_b_n1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_QO; 1.
KW ELECTRON transport; Mitochondrion; Respiratory chain; Transmembrane;

KW Heme; RNA editing. 88 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 102 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 189 IRON 2 (HEME B562 AXIAL LIGAND).
 FT METAL 203 IRON 1 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 398 AA; 44906 MW; 6FD6C5DDDCDD1DC86 CRC64;

Alignment Scores:
 Pred. No.: 5.75 Length: 398
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x CYB_WHEAT (1-398)
 QY 24 CTTTCAGTTTCTTCTCTCTGTTT 1
 DB 361 ProServAlPhePheLeuPhe 368

RESULT 4
 VE5_BPVI
 ID VE5_BPVI STANDARD; PRT; 44 AA.
 AC P06928;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE E5 protein.
 GN E5.
 OS Bovine papillomavirus type 1.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=10559;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83012974; PubMed=6289124;
 RA Chen E.Y., Howley P.M., Levinson A.D., Seeburg P.H.;
 RT "The primary structure and genetic organization of the bovine
 RT papillomavirus type 1 genome.";
 RL Nature 299:529-534(1982).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Groff D.E., Mitra R., Lancaster W.D.;
 RL Submitted (MAY-1988) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP COMPARATIVE ANALYSIS OF HUMAN TYPE 1A & BOVINE TYPE 1 GENOMES.
 RX MEDLINE=83189357; PubMed=6302319;
 RA Danos O., Engel L.W., Chen E.Y., Yaniv M., Howley P.M.;
 RT "Comparative analysis of the human type 1a and bovine type 1
 RT papillomavirus genomes.";
 RL J. Virol. 46:557-566(1983).
 RN [4]
 RP IDENTIFICATION OF PROTEIN.
 RX MEDLINE=86261794; PubMed=3014660;
 RA Schlegel R., Wade-Glass M., Rabson M.S., Yang Y.C.;
 RT "The E5 transforming gene of bovine papillomavirus encodes a small,
 RT hydrophobic polypeptide.";
 RL Science 233:464-467(1986).
 RN [5]
 RP MUTAGENESIS.
 RX MEDLINE=90012360; PubMed=2552177;
 RA Rawls J.A., Loewenstein P.M., Green M.;
 RT "Mutational analysis of bovine papillomavirus type 1 E5 peptide
 RT domains involved in induction of cellular DNA synthesis.";
 RL J. Virol. 63:4962-4964(1989).
 RN [6]
 RP INTERACTION WITH A CELLULAR PROTEIN.
 RX MEDLINE=90107936; PubMed=1688529;
 RA Goldstein D.J., Schlegel R.;
 RT "The E5 oncoprotein of bovine papillomavirus binds to a 16 kd
 RT cellular protein.";
 RL EMBO J. 9:137-146(1990).

[7]
 RN ACTIVATION OF PDGF RECEPTOR.
 RX MEDLINE=91184130; PubMed=1849073;
 RA Petti L., Nilson L.A., Dimaio D.;
 RT "Activation of the platelet-derived growth factor receptor by the
 RT bovine papillomavirus E5 transforming protein.";
 RL EMBO J. 10:845-855(1991).
 CC -1- FUNCTION: E5 CAN INDUCE CELLULAR DNA SYNTHESIS. IT SEEMS TO
 CC INTERACT WITH A 16 kDa CELLULAR PROTEIN. E5 SEEMS TO ACTIVATE
 CC THE PDGF RECEPTOR.
 CC -1- SUBUNIT: DIMER OR MONOMER.
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; X02346; CAB46513.1; -
 DR EMBL; M20219; AAA66838.1; -
 DR PIR; B18151; W5WLEB.
 DR PIR; F18151; W5WLB2.
 DR PIR; S12366; S12366.
 KW Early protein; Oncogene; Membrane.
 FT DOMAIN 32 44
 SQ SEQUENCE 44 AA; 5210 MW; 0E278099E06F2956 CRC64;

Alignment Scores:
 Pred. No.: 78.8 Length: 44
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.22% Indels: 0
 DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x VE5_BPVI (1-44)
 QY 278 CTGCTCTTTCTCTCTCTCTG 258
 DB 20 LeuLeuLeuPheLeuLeuLeu 26

RESULT 5
 YJPI_YEAST
 ID YJPI_YEAST STANDARD; PRT; 121 AA.
 AC P14359;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 13.7 kDa protein in INO1-IDS2 intergenic region.
 GN YJL151C OR J0630.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89093118; PubMed=2642902;
 RA Dean-Johnson M., Henry S.A.;
 RT "Biosynthesis of inositol in yeast. Primary structure of
 RT myo-inositol-1-phosphate synthase (EC 5.5.1.4) and functional
 RT analysis of its structural gene, the INO1 locus.";
 RL J. Biol. Chem. 264:1274-1283(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=96408771; PubMed=8813765;
 RA Katsoulou C., Tzermia M., Tavernarakis N., Alexandraki D.;
 RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast
 RT chromosome X reveals 14 known genes and 13 new open reading frames
 RT including homologues of genes clustered on the right arm of

chromosome XI.";
RL Yeast 12:787-797(1996).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -|- SIMILARITY: BELONGS TO THE UPF0057 (PMP3) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; J04453; AAA66309.1; ALT INIT.
DR EMBL; Z49427; CAA89447.1; ALT INIT.
DR EMBL; X87371; CAA60804.1; ALT_INIT.
DR PIR; A32209; A32209.
DR PIR; B30902; B30902.
DR SGD; S0003687; YJL151C.
DR InterPro; IPR000612; UPF0057.
DR Pfam; PF01679; UPF0057; 1.
DR PROSITE; PS01309; UPF0057; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 5 25
FT TRANSMEM 37 57 POTENTIAL.
SQ SEQUENCE 121 AA; 13656 MW; 3A367FF4AB00952F CRC64;

Alignment Scores:
Pred. No.: 69.1 Length: 121
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x XJPI_YEAST (1-121)

QY 278 CTGCTCTTTCTTCCTCTG 258
Db 39 LeuLeuLeuPheLeuLeu 45
|||||
RESULT 6
Y007_UREPA
ID Y007_UREPA STANDARD; PRT; 186 AA.
AC Q9PRD8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein U0007.
GN U0007.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen ureaplasma
RT urealyticum".
RL Nature 407:757-762(2000).
CC
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -|- SIMILARITY: TO U.PARVUM U0008, U0041 AND U0042.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; AE002100; AAF30412.1; -
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 80 100 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
SQ SEQUENCE 186 AA; 20203 MW; 5E06ADD8B7B58CAB CRC64;

Alignment Scores:
Pred. No.: 65.4 Length: 186
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x Y007_UREPA (1-186)

QY 281 AGCTGCTCTTTCTTCCTCTT 261
Db 78 SerLeuLeuPheLeuLeu 84
|||||
RESULT 7
YD48_AQUAE
ID YD48_AQUAE STANDARD; PRT; 189 AA.
AC O67363;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_1348.
GN AQ_1348.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:353-358(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE000736; AAC07334.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 189 AA; 21788 MW; A70F714263221FFE CRC64;

Alignment Scores:
Pred. No.: 65.3 Length: 189
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x YD48_AQUAE (1-189)

QY 281 AGCTGCTCTTTCTTCCTCTT 261
Db 29 SerLeuLeuPheLeuLeu 35
|||||
RESULT 8
YEL9_YEAST

AC YEL9_YEAST STANDARD; PRT; 196 AA.
ID PA0018;
CC
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 22.4 kDa protein in GAL83-YPT8 intergenic region.
GN YER029C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman A., Hartzell G., Hunkle-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SNRNP SMB/SMN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18778; AAB64562.1; -
CC HSSP; P14678; 1D3B.
DR SGD; S0000831; YER029C.
DR InterPro: IPR001163; snRNP_Sm.
DR Pfam: PF01423; Sm: 1.
KW Hypothetical protein; Nuclear protein; Ribonucleoprotein; RNA-binding.
SQ SEQUENCE 196 AA; 22379 MW; 0A1078D5AA241430 CRC64;

Alignment Scores:
Pred. No.: 65 Length: 196
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x YEL9_YEAST (1-196)
QY 261 AAGAGGAAAGAGCAGCGCT 281
Db 113 LysylsLysLysLysGlnAla 119
|||||
RESULT 9
Y115_NPVOP
ID Y115_NPVOP STANDARD; PRT; 205 AA.
AC O10354;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 22.1 kDa protein (ORF115).
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OPMVV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";

RL Virology 229:381-399(1997).
CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U75930; AAC59114.1; -
CC KW Hypothetical protein.
SQ SEQUENCE 205 AA; 22110 MW; A855D9AED3FF4A08 CRC64;

Alignment Scores:
Pred. No.: 64.6 Length: 205
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x Y115_NPVOP (1-205)
QY 275 CTCTTTTCTCTCTCTCTGTT 255
Db 7 LeuLeuPheLeuLeuLeuVal 13
|||||
RESULT 10
RS3_AERPE
ID RS3_AERPE STANDARD; PRT; 246 AA.
AC Q91F78;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S3P.
GN RPS3P OR APE0363.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -1- SIMILARITY: BELONGS TO THE S3P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC -1- SIMILARITY: CONTAINS 1 KH TYPE-2 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000059; BAA79318.1; -
DR InterPro: IPR004044; KH_TYPE_2.
DR InterPro: IPR004087; KH_dom.
DR InterPro: IPR004086; KH_type_1.
DR InterPro: IPR001351; Ribosomal_S3.
DR Pfam: PF00013; KH-domain; 1.
DR Pfam: PF00189; Ribosomal_S3_C; 1.

DR SMART: SM00322; KH: 1.
 DR TIGRFAMS; TIGR01008; Ipsc_EA; 1.
 DR PROSITE; PSS0084; KH TYPE 1; 1.
 DR PROSITE; PSS0823; KH TYPE 2; 1.
 DR PROSITE; PSS0548; RIBOSOMAL_S3; 1.
 KW Ribosomal protein; Complete proteome.
 FT DOMAIN 41 89 KH TYPE-2.
 SQ SEQUENCE 246 AA; 27749 MW; 4308CDB4854C870 CRC64;

Alignment Scores:
 Pred. No.: 63.1 Length: 246
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.22% Indels: 0
 DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x RS3_AERPE (1-246)

QY 92 TTCGAGGAAGAACTGCTTCA 72
 Db 218 LeuGlnGluThrAlaSer 224

RESULT 11

MODA_ECOLI STANDARD; PRT; 257 AA.
 AC P37329;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Molybdate-binding periplasmic protein precursor.
 GN MODA OR B0763.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=95164501; PubMed=7860583;
 RA Rech S., Deppenmeier U., Gunsalus R.P.;
 RT "Regulation of the molybdate transport operon, modABCD, of
 Escherichia coli in response to molybdate availability."
 RL J. Bacteriol. 177:1023-1029(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MC1000;
 RX MEDLINE=96151473; PubMed=8564363;
 RA Walkenhorst H.M., Hemschemeier S.K., Eichenlaub R.;
 RT "Molecular analysis of the molybdate uptake operon, modABCD, of
 Escherichia coli and modK, a regulatory gene."
 RL Microbiol. Res. 150:347-361(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
 RA Sanpei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 12.7-28.0 min region on the linkage map."
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE OF 25-31, AND CHARACTERIZATION.
 RX MEDLINE=96161991; PubMed=8576221;
 RA Rech S., Wolin C., Gunsalus R.P.;
 RT "Properties of the periplasmic ModA molybdate-binding protein of
 Escherichia coli."
 RL J. Biol. Chem. 271:2557-2562(1996).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS) OF 25-257.
 RX MEDLINE=97448671; PubMed=9302996;
 RA Hu Y., Rech S., Gunsalus R.P., Rees D.C.;
 RT "Crystal structure of the molybdate binding protein ModA."
 Nat. Struct. Biol. 4:703-707(1997).
 CC -I- FUNCTION: INVOLVED IN THE TRANSPORT OF MOLYBDENUM INTO THE CELL.
 CC BINDS MOLYBDATE WITH HIGH SPECIFICITY AND AFFINITY.
 CC -I- SUBCELLULAR LOCATION: Periplasmic.
 CC -I- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 CC PROTEIN FAMILY MODA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

EMBL: L34009; AAB00835.1;
 EMBL: U27192; AAB60171.1;
 EMBL: U27867; AAB06893.1;
 EMBL: AE000179; AAC73850.1;
 EMBL: D90715; BAA35427.1;
 PDB: 1AMF; 24-DEC-97.
 PDB: 1WOD; 17-DEC-97.
 DR SWISS-2DPAGE; P37329; COLI.
 DR EcoGene; EGI2427; modA.
 DR TIGRFAMS; TIGR01256; modA; 1.
 KW Transport; Molybdenum; Periplasmic; Signal; 3D-structure;
 KW Complete proteome.
 FT SIGNAL 1 24
 FT CHAIN 25 257
 SQ SEQUENCE 257 AA; 27364 MW; 847DB740EE0564E7 CRC64;

Alignment Scores:

Pred. No.: 62.8 Length: 257
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.07% Indels: 0
 DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x MODA_ECOLI (1-257)

QY 58 TTGGTGAAGGAAGTGAAGCA 78
 Db 182 LeuValGluArgAsnGluAla 188

RESULT 12

PLCA_HUMAN STANDARD; PRT; 283 AA.
 AC Q99943;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 1-acyl-sn-glycerol-3-phosphate acyltransferase alpha (EC 2.3.1.51) (1-
 DE AGP acyltransferase 1) (1-AGPAT 1) (Lysophosphatidic acid
 DE acyltransferase-alpha) (LPAT-alpha) (1-acylglycerol-3-phosphate O-
 DE acyltransferase 1).
 GN AGPAT1 OR G15.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 RA MEDLINE=97355682; PubMed=9212163;
 RA West J., Tompkins C.K., Balantac N., Nudelman E., Meengs B., White T.,
 RA Bursten S., Coleman J., Kumar A., Singer J.W., Leung D.W.;
 RA "Cloning and expression of two human lysophosphatidic acid
 RT acyltransferase cDNAs that enhance cytokine-induced signaling
 RT responses in cells.";
 RL DNA Cell Biol. 16:691-701(1997).
 RL [2]
 RP SEQUENCE OF 9-283 FROM N.A.
 RA Rowen L., Danks C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
 RA Banta A., Spies T., Hood L.;
 RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97419100; PubMed=9291118;
 RA Stamps A.C., Elmore M.A., Hill K., Kelly K., Makda A.A.,
 RA Finnen M.J.;
 RA "A human cDNA sequence with homology to non-mammalian lysophosphatidic
 RT acid acyltransferases.";
 RL Biochem. J. 326:455-461(1997).
 RL [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98129822; PubMed=9461603;
 RA Aguado B., Campbell R.D.;
 RA "Characterization of a human lysophosphatidic acid acyltransferase
 RT that is encoded by a gene located in the class III region of the
 RT human major histocompatibility complex.";
 RL J. Biol. Chem. 273:4096-4105(1998).
 RL [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Lung;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CONVERTS LYSPHOSPHATIDIC ACID (LPA) INTO PHOSPHATIDIC
 CC ACID BY INCORPORATING ACYL MOIETY AT THE 2 POSITION.
 CC -1- CATALYTIC ACTIVITY: Acyl-CoA + 1-acyl-sn-glycerol 3-phosphate =
 CC CoA + 1,2-diacyl-sn-glycerol 3-phosphate.
 CC -1- PATHWAY: De novo phospholipid biosynthesis; second step.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE 1-ACYL-SN-GLYCEROL-3-PHOSPHATE
 CC ACYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U56417; AAB58775.1; -;
 DR EMBL; U89336; AAB47493.1; -;
 DR EMBL; U75971; AAB96378.1; -;
 DR EMBL; Y09565; CAA70758.1; -;
 DR EMBL; BC002402; AAH02402.1; -;
 DR EMBL; BC003007; AAH03007.1; -;
 DR EMBL; BC004310; AAH04310.1; -;
 DR Genew; HGNC:324; AGPAT1.
 DR MIM; 603099; -;
 DR InterPro; IPR004552; AGP_acyltnr.
 DR InterPro; IPR002123; Acyltransferase.
 DR Pfam; PF01553; Acyltransferase; 1.
 DR TIGRfams; TIGR00530; AGP_acyltnr; 1.
 DR Phospholipid biosynthesis; Transferase; Acyltransferase;
 KW Transmembrane.
 FT TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 128 148 POTENTIAL.
 SQ SEQUENCE 283 AA; 31716 MW; 71F3207259747C68 CRC64;

Alignment Scores: 62 Length: 283
 Pred. No.: 7.00 Matches: 7
 Score: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Indels: 0
 DB: 1 Gaps: 0
 US-10-060-830-4 (1-300) x PLCA_HUMAN (1-283)
 QY 278 CTGCTTCTTCTCTCTCTG 258
 Db 12 LeuLeuLeuPheLeuLeuLeu 18
 RESULT 13
 PRMA_CLOAB STANDARD; PRT; 311 AA.
 ID PRMA_CLOAB STANDARD; PRT; 311 AA.
 AC P45558;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable methyltransferase (EC 2.1.1.-).
 GN PRMA OR CAC1284.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RA "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum.";
 RL J. Bacteriol. 183:4823-4838(2001).
 RN [2]
 RP SEQUENCE OF 1-298 FROM N.A.
 RX MEDLINE=94123950; PubMed=7507453;
 RA Behrens S., Narberhaus F., Bahl H.;
 RA "Cloning, nucleotide sequence and structural analysis of the
 RT Clostridium acetobutylicum dnaJ gene.";
 RL FEMS Microbiol. Lett. 114:53-60(1993).
 RN [3]
 RP REVISIONS.
 RA Behrens S.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- INDUCTION: BY HEAT SHOCK.
 CC -1- SIMILARITY: STRONG, TO ENTEROBACTERIAL RIBOSOMAL PROTEIN L11
 CC METHYLTRANSFERASE (PRMA).
 CC -1- SIMILARITY: TO OTHER METHYLTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB007641; AAK79255.1; -;
 DR EMBL; X69050; CAA48793.1; -;
 DR InterPro; IPR004498; PRMA.
 DR InterPro; IPR000051; SAM_bind.
 DR TIGRfams; TIGR000406; PRMA; 1.
 KW Heat shock; Transferase; Methyltransferase; Complete proteome.
 SQ SEQUENCE 311 AA; 35028 MW; 2D71786B3C974922 CRC64;

Alignment Scores: 61.2 Length: 311
 Pred. No.:

Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x PRMA_CLOAB (1-311)

QY 123 GCTATAGCAGCAGCAACTTA 143
Db 191 AlAlIleAlaAlaAlaLysLeu 197

RESULT 14

VAOD_HUMAN STANDARD; PRT; 351 AA.
AC P12953; Q02547;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vacuolar ATP synthase subunit d (EC 3.6.3.14) (V-ATPase d subunit)
DE (Vacuolar proton pump d subunit) (V-ATPase AC39 subunit) (V-ATPase 40
DE kDa accessory protein) (P39) (32 kDa accessory protein).
GN ATP6V0D1 OR ATP6D OR VPATPD.
OS Homo sapiens (Human), and
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCB1_TaxID-9606, 9913;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-Human;
RX MEDLINE-20568713; PubMed-11118322;
RA Agarwal A.K., White P.C.;
RT "Structure of the VPAATPD gene encoding subunit D of the human vacuolar
RT proton ATPase.";
RL Biochem. Biophys. Res. Commun. 279:543-547(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-Human;
RX MEDLINE-94071935; PubMed-8250920;
RA van Hille B., Vanek M., Richener H., Green J.R., Bilbe G.;
RT "Cloning and tissue distribution of subunits C, D, and E of the human
RT vacuolar H(+)ATPase.";
RL Biochem. Biophys. Res. Commun. 197:15-21(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-Human; TISSUE-Brain;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 264-351 FROM N.A.
RC SPECIES-Human;
RA Bhat K.S.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC SPECIES-Bovine; TISSUE-Adrenal medulla;
RX MEDLINE-89034297; PubMed-2903164;
RA Wang S.-Y., Moriyama Y., Mandel M., Hulmes J.D., Pan Y.-C.E.,
RA Danho W., Nelson H., Nelson N.;
RT "Cloning of cDNA encoding a 32-kDa protein. An accessory polypeptide
RT of the H-ATPase from chromaffin granules.";
RL J. Biol. Chem. 263:17638-17641(1988).
CC -!- FUNCTION: SUBUNIT OF THE INTEGRAL MEMBRANE V0 COMPLEX OF VACUOLAR
CC ATPASE. VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY OF
CC INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS, THUS PROVIDING
CC MOST OF THE ENERGY REQUIRED FOR TRANSPORT PROCESSES IN THE
CC VACUOLAR SYSTEM.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC H(+)(Out).
CC -!- SUBUNIT: V-ATPase is a heteromultimeric enzyme composed of a
CC peripheral catalytic V1 complex (components A to H) attached to
CC an integral membrane V0 proton pore complex (components: a, c, c',

c' and d).
-!- P-TM: THE N-TERMINUS IS BLOCKED.
-!- SIMILARITY: BELONGS TO THE V-ATPASE V0D/AC39 SUBUNIT FAMILY.
-!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N-TERMINAL
SECTION DUE TO FRAMESHIFTS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).

EMBL; X71490; CAA50591.1; ALT_FRAME.
EMBL; BC008861; AA008861.1;
EMBL; L05087; AAC15852.1;
EMBL; J04204; AAA64520.1;
PIR; JN0908; JN0908.
PIR; S34132; S34132.
PIR; A32123; A32123.
Genew; HGNC:13724; ATP6V0D1.
MIM; 607028;
InterPro: IPR002843; ATPSynt_AC39sub.
Pfam; PF01992; V-ATP-synt_AC39; 1.
KW Hydrolyase; Hydrogen ion transport.
FT CONFLICT 27 27 V -> E (IN REF. 1).
FT CONFLICT 66 66 T -> A (IN REF. 1).
FT CONFLICT 266 267 NV -> KL (IN REF. 1).
SQ SEQUENCE 351 AA; 40329 MW; A720F8A87511203C CRC64;

Alignment Scores:
Pred. No.: 60.3 Length: 351
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x VAOD_HUMAN (1-351)
QY 227 GGTACACTTCATCAAGATCT 247
Db 115 GlyThrLeuHisGlnArgSer 121

RESULT 15
VAOD_MOUSE
ID VAOD_MOUSE STANDARD; PRT; 351 AA.
AC P51863; Q9QWJ2;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vacuolar ATP synthase subunit d (EC 3.6.3.14) (V-ATPase d subunit)
DE (Vacuolar proton pump d subunit) (V-ATPase AC39 subunit) (V-ATPase 40
DE kDa accessory protein) (P39) (Physophillin).
GN ATP6V0D1 OR ATP6D.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID-10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Howell M.L., Dean G.E.;
RT "cDNA sequences for mouse vacuolar ATPase subunits.";
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-Brain;
RA Carrion-Vazquez M., Fernandez A., Chowen J., Nieto-Sampedro M.;
RT "Cloning and expression of AC39/physophillin in rodent brain.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SUBUNIT OF THE INTEGRAL MEMBRANE V0 COMPLEX OF VACUOLAR
CC ATPASE. VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY OF

```

CC INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS, THUS PROVIDING
CC MOST OF THE ENERGY REQUIRED FOR TRANSPORT PROCESSES IN THE
CC VACUOLAR SYSTEM.
CC -|- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) -> ADP + phosphate +
CC H(+)(Out).
CC -|- SUBUNIT: V-ATPase is a heteromultimeric enzyme composed of a
CC peripheral catalytic VI complex (components A to H) attached to
CC an integral membrane V0 proton pore complex (components: a, c, c',
CC c'', and d).
CC -|- SIMILARITY: BELONGS TO THE V-ATPASE V0D/AC39 SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U13840; AAC83085.1; -
DR EMBL; U21549; AAA92288.1; -
DR MGD; MGI:1201778; Atp6d.
DR InterPro: IPR002843; ATPsynt_AC39sub.
DR Pfam: PF01992; VATP-synt_AC39; 1.
KW Hydrolase; Hydrogen ion transport.
FT CONFLICT 23 23 L -> M (IN REF. 1).
SQ SEQUENCE 351 AA; 40301 MW; 62CDF67B982124C9 CRC64;

Alignment Scores:
Pred. No.: 60.3 Length: 351
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x VA0D_MOUSE (1-351)
QY 227 GGTACACTTCATCAAGATCT 247
Db 115 GlyThrLeuHisGlnArgSer 121

```

Search completed: January 21, 2003, 14:49:55
 Job time : 12.2126 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:43:31 ; Search time 30.9449 Seconds
(without alignments)
3995.112 Million cell updates

Title: US-10-060-830-4

```

Perfect score: 99
Sequence: 1 aaacgaagaaqaaaaaactg.....tatgcagacctagatccta 300

```

Scoring table:	OLIGO	Xgapop	60.0	Xgapext	60.0
		Ygapop	60.0	Ygapext	60.0
		Fgapop	6.0	Fgapext	7.0
		Delop	6.0	Delext	7.0

Searched: 671580 seqs, 206047115 residues

Word size:

Total number of hits satisfying chosen parameters: 1343044

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

```

-MODEL=framet_n2p_spool_032702_1498/app_query_fasta_1.1109
-O/cn2_1/us10060830/runat_16012003_092702_1498
DB=SPREMBL_21_QFMT=fastan_SUFFIX=olin2p.rspt_MINMATCH=0.1-LOOCL=0
-LOOPEXT=0-UNITS=bits-START=1-END=-1-WATRIX=oligo-TRANS=human40.cdi
-LIST=45-DLOCALCN=200-THR_SCORE=quality-THR_MIN=1-ALIGN=15-MODE=LOCAL
-OUTFMT=ptc-NORM=ext-HEAPSIZE=500-MINLEN=0-MAXLEN=2000000000
-USER=us10060830_@CN_1_1_57_@runat_16012003_092702_1498-NCPU=6-ICPU=3
-NO_XLPXY-NO_MMAP-LARGEQUERY-NEG_SCORES=0-WAIT-LONGLOG-DEV_TIMEOUT=120
-WARN_TIMEOUT=30-THREADS=1-XGAPOP=60-XGAPEXT=60-FGAPOP=6-FGAPEXT=7
-YGAPOP=60-YGAPEXT=60-DELOP=6-DELEXT=7

```

```
Database :
SPREMBL_21.*
1:  sp.archea.*
2:  sp.bacteria.*
3:  sp.fungi.*
4:  sp.human.*
5:  sp.invertebrate.*
6:  sp.mammal.*
7:  sp.mhc.*
8:  sp.organella.*
9:  sp.phage.*
10: sp.plant.*
11: sp.rodent.*
12: sp.virus.*
13: sp.vertebrate.*
14: sp.unclassified.*
15: sp.rviro.*
16: sp.bacteriap.*
17: sp.archeap.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	100.0	775	4 Q96PD2	Q96pd2 homo sapien

DR EMBL: AF387547; AAL30178.1; -;
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR00421; FA58_C.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00754; F5_P8_type_C; 1.
 DR PROSITE: PS01180; CUB; 1.
 KW DNA-binding; Transcription regulation.
 SQ SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;

Alignment Scores:
 Pred. No.: 1.71e-94 Length: 775
 Score: 99.00 Matches: 99
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-060-830-4 (1-300) x Q96PD2 (1-775)

QY 2 AACAGAAAGAAAAAAGTGAAGGCACCTATGACTTACCTTACTGGGACCGGCGAGTGG 61
 Db 556 AsnArgLysLysLysThrGluGlyThrTyraSpLeuProTyrTipAspArgAlaGlyTrp 575
 QY 62 TGGAAAGGAATGAAGCAGTCTTCTCCTCAAAAGCAGTGGACCATGAGAAACCCAGTT 121
 Db 576 TrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProVal 595
 QY 122 CGCTATAGCAGCAGCAAGTAAATCACTGAGTCAAGAGAAAGTCAACACAGTGTCTGCAG 181
 Db 596 ArgTyrSerSerGluValAsnHisLeuSerProArgGluValThrValLeuGln 615
 QY 182 GCTGACTCTGCAGAGTATGCTCAGCCACTGGTAGGAGAAATGTTGGTACACTTCATCAA 241
 Db 616 AlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGln 635
 QY 242 AGATCTACCTTTAAACAGAGAAAGCAAGAGCAAGTATGACAGCTAGATCCT 298
 Db 636 ArgSerThrPheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspPro 654

RESULT 3
 Q8DX2 Q9D9K5 PRELIMINARY; PRT; 251 AA.
 ID Q9D9K5 PRELIMINARY; PRT; 251 AA.
 AC Q9D9K5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 1700055P21RIK protein.
 GN ESDN OR 1700055P21RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schraml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Blake J., Boffelli D., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Brownstein M.J., Bull C., Fletcher C., Carninci P., de Bonaldo M.F.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK006805; BAB24750.1; -;
 DR MGI:1920629; Esdn.
 SQ SEQUENCE 251 AA; 27428 MW; F86E0AA15EF51AE9 CRC64;

Alignment Scores:
 Pred. No.: 3.85e-40 Length: 251
 Score: 47.00 Matches: 47
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 47.47% Indels: 0
 DB: 11 Gaps: 0

US-10-060-830-4 (1-300) x Q9D9K5 (1-251)

QY 158 AGAAGTCAACACAGTCTGCAGCTGACTCTGCAGAGTATGCTCAGCCACTGGTAGGA 217

DR EMBL: AF387547; AAL30178.1; -;
 DR InterPro: IPR000859; CUB_domain.
 DR InterPro: IPR00421; FA58_C.
 DR InterPro: IPR004043; LCCL_dom.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00754; F5_P8_type_C; 1.
 DR PROSITE: PS01180; CUB; 1.
 KW DNA-binding; Transcription regulation.
 SQ SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;

Alignment Scores:
 Pred. No.: 1.71e-94 Length: 775
 Score: 99.00 Matches: 99
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-060-830-4 (1-300) x Q96PD2 (1-775)

QY 2 AACAGAAAGAAAAAAGTGAAGGCACCTATGACTTACCTTACTGGGACCGGCGAGTGG 61
 Db 556 AsnArgLysLysLysThrGluGlyThrTyraSpLeuProTyrTipAspArgAlaGlyTrp 575
 QY 62 TGGAAAGGAATGAAGCAGTCTTCTCCTCAAAAGCAGTGGACCATGAGAAACCCAGTT 121
 Db 576 TrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProVal 595
 QY 122 CGCTATAGCAGCAGCAAGTAAATCACTGAGTCAAGAGAAAGTCAACACAGTGTCTGCAG 181
 Db 596 ArgTyrSerSerGluValAsnHisLeuSerProArgGluValThrValLeuGln 615
 QY 182 GCTGACTCTGCAGAGTATGCTCAGCCACTGGTAGGAGAAATGTTGGTACACTTCATCAA 241
 Db 616 AlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGln 635
 QY 242 AGATCTACCTTTAAACAGAGAAAGCAAGAGCAAGTATGACAGCTAGATCCT 298
 Db 636 ArgSerThrPheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspPro 654

RESULT 2
 Q8DX2 Q9D9K5 PRELIMINARY; PRT; 775 AA.
 ID Q8DX2 PRELIMINARY; PRT; 775 AA.
 AC Q8DX2;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE CLCP1.
 GN CLCP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Koshikawa K., Osada H., Kozaki K., Konishi H., Masuda A.,
 RA Tatematsu Y., Mitsudomi T., Nakao A., Takahashi T.;
 RT "Significant up-regulation of a novel gene, CLCP1, in a highly
 RT metastatic lung cancer subline as well as in lung cancers in vivo."
 RL Oncogene 0:0-0(2002).
 DR EMBL: AB073146; BAB91138.1; -;
 SQ SEQUENCE 775 AA; 85073 MW; 6D07223B21BE5A42 CRC64;

Alignment Scores:
 Pred. No.: 1.71e-94 Length: 775
 Score: 99.00 Matches: 99
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-060-830-4 (1-300) x Q8DX2 (1-775)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 21, 2003, 09:58:26 ; Search time 4536.98 Seconds
(without alignments)
4188.719 Million cell updates/sec

Title: US-10-060-830-1114
Perfect score: 3470
Sequence: 1 MFLFLLLLVLLLEEDAGA.....TOEVSAGRGDCDVFKEIL 553

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10060830/runat_16012003_091559_29008/app_query.fasta_1.1742
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=oct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.cgn_1_1_633_@runat_16012003_091559_29008 -NCPU=6 -ICPU=3
-NO_XUPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pln:*
35: em_hgt_rod:*
36: em_hgt_mam:*
37: em_hgt_vrt:*
38: em_sy:*
39: em_hgtgo_hum:*
40: em_hgtgo_mus:*
41: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3422	98.6	2939	9	AF387547	AF387547 Homo sapi
2	3422	98.6	6093	9	AB073146	AB073146 Homo sapi
3	3405	98.1	3858	9	BC029658	BC029658 Homo sapi
4	2984.5	86.0	2339	10	AF387549	AF387549 Rattus no
5	2950.5	85.0	2642	10	AF387548	AF387548 Mus muscu
6	2226	64.1	2020	9	AK056350	AK056350 Homo sapi
7	2126.5	61.3	1907	6	AK405795	AK405795 Sequence
8	1712.5	49.4	1388	9	HUMCUB1	D29810 Human mRNA
9	922.5	26.6	192247	2	AC013497	AC013497 Homo sapi
10	922.5	26.6	193623	9	AC106728	AC106728 Homo sapi
11	870	25.1	2547	9	AK095973	AK095973 Homo sapi
12	852.5	24.6	203102	2	AC126975	AC126975 Rattus no
13	825	23.8	146515	2	AC025661	AC025661 Homo sapi
14	756	21.8	1464	6	AX118816	AX118816 Sequence
15	756	21.8	1620	6	AX118820	AX118820 Sequence
16	756	21.8	1761	6	AX118818	AX118818 Sequence
17	756	21.8	1768	6	AX118822	AX118822 Sequence
18	429	12.4	192247	2	AC013497	AC013497 Homo sapi
19	414	11.9	2846	10	BC026771	BC026771 Mus muscu
20	367	10.6	6585	10	MMU52925	U52925 Mus musculu
21	362	10.4	7062	4	AF191308	AF191308 Sus scrof
22	361	10.4	6895	4	BOVFACV2A	M81441 Bos taurus
23	361	10.4	6910	4	BOVFACV	M81440 Bos taurus
24	357.5	10.3	6893	9	HUMFVA	M14335 Human coagu
25	356.5	10.3	1895	4	S80643	S80643 Bos taurus
26	356.5	10.3	1965	4	BTPAS67PT	X91895 B.taurus mR
27	355	10.2	1719	9	HSU70312	U70312 Homo sapien
28	355	10.2	2728	9	BC030828	BC030828 Homo sapi
29	355	10.2	4681	6	AX473349	AX473349 Sequence
30	353.5	10.2	6909	6	A63255	A63255 Sequence 1
31	353.5	10.2	6909	6	A63218	A63218 Sequence 1
32	353.5	10.2	6909	6	AR071379	AR071379 Sequence
33	353.5	10.2	6909	6	AX146885	AX146885 Sequence
34	353.5	10.2	6909	6	AX409628	AX409628 Sequence
35	353.5	10.2	6909	6	E29754	E29754 Method for
36	353.5	10.2	6909	9	HUMFVA	M16967 Human coagu
37	352	10.1	2303	10	AF031524	AF031524 Mus muscu
38	348	10.0	1227	4	BTBP47PRO	Y11719 Bos taurus
39	346	10.0	1270	6	AX410374	AX410374 Sequence
40	346	10.0	1270	9	S56151	S56151 HMFG-milk f
41	346	10.0	1384	6	AR081092	AR081092 Sequence
42	346	10.0	1384	6	I14823	I14823 Sequence 1
43	346	10.0	1934	6	AR081093	AR081093 Sequence
44	346	10.0	1934	9	HSU58516	U58516 Human breas
45	346	10.0	1987	9	AK095908	AK095908 Homo sapi

ALIGNMENTS

RESULT 1

AF387547	2939 bp	mRNA	linear	PRI 11-NOV-2001
LOCUS	Homo sapiens endothelial and smooth muscle cell-derived neuropilin-like protein (ESDN) mRNA, complete cds.			
DEFINITION				
ACCESSION	AF387547			
VERSION	AF387547.1	GI:16902434		
KEYWORDS				
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 2939) Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N., Matsumori,A., Sasayama,S., Honjo,T. and Tashiro.K. ESDN, a novel neuropilin-like membrane protein cloned from vascular cells with the longest secretory signal sequence among eukaryotes, is up-regulated after vascular injury J. Biol. Chem. 276 (36), 34105-34114 (2001)			
AUTHORS	21424027			
TITLE	11447234			
JOURNAL	2 (bases 1 to 2939) Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N., Sasayama,S., Matsumori,A., Honjo,T. and Tashiro,K. Direct Submission Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan			
MEDLINE				
PUBMED				
REFERENCE	Location/Qualifiers			
AUTHORS	1..2939 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="3"			
TITLE	1..2939 /gene="ESDN"			
JOURNAL	12..2339 /gene="ESDN"			
FEATURES	/codon_start=1 /product="endothelial and smooth muscle cell-derived neuropilin-like protein" /protein_id="AAL30178.1" /db_xref="GI:16902435" /translation="MASRVARRRCPQCPQVRAAAAPAWAALPLSRSLPPCSNSSS FSFPLGLLVLLLEDAAGCGGCHTVLGPSEGTLSINYPTQYPSNVCEIE IRVMGERVIRKFGEFDIEDSDCHFNYLIYNGISVRTEIGKYCGLQMHNHSIE KGNETLPHSGTHVSGRFLASVYIDKODLTCLDTASNLFPEPFSKYPACGLPLP FAETSGTTPHGYSRDSPLWAGVHVNSNTLGQISVTISKIPYVESLIANNVTSV VGHLSTLFTFKYSGCYTLCMESGYIADPQITASSVLENTDHTGOENSKPKRKARLK KPQPMPGAATDEYIQMLIDLNEKKTLGITTSCTVEHNIYVSATRILI SDDGQKW TVYREPEVEDQIKFOGNKHQDVNRNNLPPIIARFIRVNPTQOQR IAKMELLGQV FASPDRKITQTTPPNRSNDLKNTPAPRIAGRAPKFTQPLQRSNNEFPQAETEOT TKSPDIRNTLVNLVDLAALVPVLVWLTTLLILVCWAHWNRKKKTGETGVDP LPYWDRAWGWMKQFLPAKAVDHEETPVYSSSEVNHLSPREVTTVLOADSAYEAP LVGGIVGTHQRTFFREEKEAGYADLDYPNSPGQEVYHAYAEPLPITGPEYATPPII MDMSGHGTVSGQPSFTKATGNQPPPLVGTINTLLSRDSCSSAQYDITPAAGRP GLPADBLVTVQPSQTEVSGARDGEDVFKEIL"			
BASE COUNT	825 a 675 c 657 g 782 t			
ORIGIN				
Alignment Scores:				
Pred. No.:	2.68e-225	Length:	2939	
Score:	3422.00	Matches:	653	
Percent Similarity:	89.57%	Conservative:	0	
Best Local Similarity:	89.57%	Mismatch:	0	
Query Match:	98.62%	Indels:	76	
DB:	9	Gaps:	1	
US-10-060-830-1114 (1-653) x AF387547 (1-2939)				
Oy 1 MetProLeuPeLeuLeuLeuLeuValLeuLeuLeuLeuGluAspAlaGlyAla 20				
Db 150 AFGCCCTGTCCTCCTCCTACTTGCTGCCTCCTGCTCGAGGACGCTGAGGCC 209				
Oy 21 GlnGln----- 22				

Dbb 1290 CGTTTTATTAGACTGAATCTTACCCAAATGCGCAGCAGAAAATTCGCATGAAATGGAGCTG 1349
QY 325 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProProPro 344
|||||
Db 1350 CTCGGATGTCAGTTATTCTTAAAGTCGCTCTCCAAATCTTACTCAACCTCCACCTCTCT 1409
QY 345 ArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla 364
|||||
Db 1410 CGGAACAGCAATGACCTCAAAACACTACAGCCCTCCAAAATAGCCAAAGTGGTGGC 1469
QY 365 ProLysPheThrGlnProLeuGlnProArgSerSerAsnGlnPheProAlaGlnThrGlu 384
|||||
Db 1470 CCAAATTTACGACCACTACAACTCGCAGTACGATGAATTCCTCCACAGACAGAA 1529
QY 385 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 404
|||||
Db 1530 CAAACAACTGCGCAGCTGATATCAAAATACCTACCGTAACCTCAATGTAACCAAGAT 1589
QY 405 ValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 424
|||||
Db 1590 GTAGCGCTGGCTGACGTTCTGTCCCTGTGCTGGTCAATGCTCTACTCTCATCTTC 1649
QY 425 IleLeuValCysAlaThrHisTrpArgAsnArgLysLysThrGluGlyThrTrpAsp 444
|||||
Db 1650 ATATTAGTGTGTGGCTGGAGCACTGGAGAACAGAAAACAACTCAAGCCACCTATGAC 1709
QY 445 LeuProTrpTrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLys 464
|||||
Db 1710 TTACCTTACTGGGACCGGCGAGTTGGTGGAAAGGAATGAAGCAGTTCCTCTCTGCAAA 1769
QY 465 AlaValAspHisAlaLeuThrProValArgTrpSerSerSerClnuValAsnHisLeuSer 484
|||||
Db 1770 GCAGTGGACCATGAGGAACCCCGATCTGCTATGAGCAGGCGAAGTTAATCACCTGAGT 1829
QY 485 ProArgGluValThrThrValLeuGlnAlaAspSerAlaGluThrAlaGlnProLeuVal 504
|||||
Db 1830 CCAAGAGAAGTCACACAGCTGCTGAGGCTGACTCTGCAGAGATGCTCAGCCACTGGTA 1889
QY 505 GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGlu 524
|||||
Db 1890 GGAGGAATTCGTTGGTACATCTCATCAAGATCTACCTTTAAACCAAGAGGAAAGAA 1949
QY 525 AlaGlyThrAlaAspLeuAspProTrpAsnSerProGlyGlnGluValThrHisAlaTrp 544
|||||
Db 1950 GCAGGCTATGCAGACCTAGATCTTCAACTCACCAGGCGAGGAGTTTATCATGCTAT 2009
QY 545 AlaGluProLeuProIleThrGlyProGluThrAlaThrProIleMetAspMetSer 564
|||||
Db 2010 GCTGAACCACTCCCAATTACGGGCTGAGTATGCAACCCCAATCATGACATGTCATCA 2069
QY 565 GlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsn 584
|||||
Db 2070 GGGCACCCCACTTCACTTGGTGCAGCCCTCCACATCCACTTTCAAGGCTACGGGGAAC 2129
QY 585 GlnProProLeuValGlyThrThrAsnThrLeuLeuSerArgThrAspSerCysSer 604
|||||
Db 2130 CAACCTCCCCCACTAGTGGGAATTTCAATATACACTTCTCTCCAGGACTGACACCTGCTCC 2189
QY 605 SerAlaGlnAlaGlnThrAspThrProLysAlaGlyLysProGlyLeuProAlaProAsp 624
|||||
Db 2190 TCAGCCAGCCCGCAGTATATATCCCGAAGCTGGGAGCCAGGTCTACCTGCCCCAGAC 2249
QY 625 GluLeuValThrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGly 644
|||||
Db 2250 GAATGGTGTACAGGTGCCACAGACACACAGAGAGATATCAGGACGAGGAGGATGGG 2309
QY 645 GluCysAspValPheLysGluIleLeu 653
|||||
Db 2310 GAATGTGATGTTTAAAGAAATCCCT 2336
RESULT 2
AB073146 6093 bp mRNA linear PRI 26-APR-2002
LOCUS

DEFINITION Homo sapiens mRNA for CLCP1, complete cds.
ACCESSION AB073146
VERSION AB073146.1 GI:20330503
KEYWORDS
SOURCE Homo sapiens lung cell_line:normal lung tissue cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Koshikawa, K., Osada, H., Kozaki, K., Konishi, H., Masuda, A.,
Tatematsu, Y., Mitsudomi, T., Nakao, A. and Takahashi, T.
Significant up-regulation of a novel gene, CLCP1, in a highly
metastatic lung cancer subline as well as in lung cancers in vivo
Oncogene (2002) in press
2 (bases 1 to 6093)
Koshikawa, K., Konishi, H., Osada, H. and Takahashi, T.
Direct Submission
Submitted (16-OCT-2001) Takashi Takahashi, Aichi Cancer Center
Research Institute, Division of Molecular Oncology; Chikusa-ku,
Kanokoden 1-1, Nagoya, Aichi 464-8681, Japan
(E-mail:take@ichi.cc.jp, Tel:81-52-764-2993, Fax:81-52-764-2993)
FEATURES
Location/Qualifiers
1..6093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3q12.1-12.2"
/cell_line="normal lung tissue"
/tissue_type="lung"
1..6093
/gene="CLCP1"
364..2691
/gene="CLCP1"
/function="associated with metastatic phenotype"
/note="CUB, LCCL, coagulation factor V/VIII-homology
domains protein 1"
/codon_start=1
/evidence=experimental
/product="CLCP1"
/protein_id="BAB91138.1"
/db_xref="GI:20330504"
/translation="MASRAVVRRCPCQPOVRAAXAAPAAALPLSRSLPPCSNSS
FSNPLFLILLVLLLEDAAGQGGCGHTVLGPESGTLTSINPTQYPSNVCWE
IRVKGERRVRIKFGEDIEDSDSCHENLYRIYNGIGVSRTEIKGYCGLQMHHSIES
KGEITLLEMSGIHVSGRGLASVLDKODLLCLDTCLDTSLEPEFSKTCPCALLP
FABISCTIPHGYSRDSPLCMAGVHAGVVSNTLGGQISVSKGIPYESSLANNTSV
VHLSLSTFTKTSCTGCTGCMESGVADPQITASSVLEWTDHTGOENSWKPKARK
KCPGPAAPATDEYOMLOIDLNKEKKTGTTTGTSTVMEHNYVVSAYRILYSDGQKW
TYVREPGVEDKIFQGNKDYHODVRNLFPLIARIRVNPTQOQKIANKMELLCQ
FIPGRPKLTPPPRNSNDLNTAPPAIKRGRAPKFTQPLQPRSSNEFFPAQTEQT
TASPDIRNTTVPNTVKDVALAAVLVPVLMVLTLLILLCAMHWRNRKKTGTYD
LPYWDRAWMKMGKQFLPAKAVDHEETPVYSSEVNHLSPREVTYVLQADSAEYAP
LVGGIVGTLHORSTFKPEKGAGYADLPDYPNGQOEYHAYAEPLPITGPEYATPII
MDMSGHTTSVGOSTSTFKATCNPPPLVNTYNTLLSRDTSSCSAQAQYDTPKAGP
GLPAPDELTYQVQSTQEVSGAGRDGECDFKELL"
577..924
/gene="CLCP1"
/note="CUB domain"
934..1191
/gene="CLCP1"
/note="LCCL domain"
1234..1710
/gene="CLCP1"
/note="F5/8 type C domain"
1942..2010
/gene="CLCP1"
/note="transmembrane domain"
4264..4269
/gene="CLCP1"
5873..5878
/gene="CLCP1"
6052..6057
/gene="CLCP1"

BASE COUNT	1671 a	1258 c	1289 g	1874 t	1 others
ORIGIN					

Alignment Scores:

Pred. No.:	6.6e-225	Length:	6093
Score:	3422.00	Matches:	653
Percent Similarity:	89.57%	Conservative:	0
Best Local Similarity:	89.57%	Mismatches:	0
Query Match:	98.62%	Indels:	76
DB:	9	Gaps:	1

US-10-060-830-1114 (1-653) x AB073146 (1-6093)

Qy	1	MetProLeuPheLeuLeuLeuLeuValLeuLeuLeuLeuLeuLeuGluLeuAspAlaGlyAla	20
Db	502	ATGCCTCTGTCTCCCTGCTCTTACTTGTCTGCTCTGCTCTGCTCGAGACGCTGGAGCC	561
Qy	21	GlnGln-----	22
Db	562	CAGCAAGGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAAACCTTACATCC	621
Qy	22	-----	22
Db	622	ATAAACTACCACAGACCTATCCCAACAGCACTGTTTGTGAATGGGAGATCCGCTGTAAAG	681
Qy	22	-----	22
Db	682	ATGGGAGAGAGATTGCGATCAAAATTTGGTGACTTTTGACATTGAAGATTCTGATCTTGT	741
Qy	23	-----GlyLys 24	
Db	742	CACTTTAATTACTTGAAGATTTATAATGGAATTTGGAGTCAGCAGAACTCAAAATAGCGAAA	801
Qy	25	TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr	44
Db	802	TACTGTGCTCGGGTGTGCAATGAACCATTTCAATTTGAATCAAAAGGCAATGAAATCAC	861
Qy	45	LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal	64
Db	862	TTGCTGTTTCATGAGTGGNAATCCATGTTCTGGACGCGGATTTTGGCCTCATACTCTGTT	921
Qy	65	IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPhelLeuGluProGlu	84
Db	922	ATAGATAAACAAAGATCTAATTTACTTGTTTGGACACTGCATCCAAATTTTGGAACTGAG	981
Qy	85	PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIle	104
Db	982	TTCAAGTAAGTACTGCCACAGCTGGTGTCTGCTTCCTTTTGTGAGATATCTGGAACAAT	1041
Qy	105	ProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValVal	124
Db	1042	CCTCATGTGATATAGAGATTCCTCGGCATTTGTGCATGGCTGGTGTGTCATCGAGAGTAGT	1101
Qy	125	SerAsnThrLeuGlyGlyGlnIleSerValIleSerLysGlyIleProTyrTyrGlu	144
Db	1102	TCAAAACAGTTGGCGCGCAATACAGTGTGTGAATTAGTAAGGTATCCCTTATTATGAA	1161
Qy	145	SerSerLeuAlaAsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr	164
Db	1162	AGTCTTTGGCTAACCAACGTCACATCTGTGGTGGGACACTTATCTACAAAGCTCTTTTACA	1221
Qy	165	PhelysThrSerGlyCystyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspPro	184
Db	1222	TTTAAGACAAGTGGATGTATTGGAACACTGGGGAGTGGAGTCTGGTGTGATCGCGGATCCT	1281
Qy	185	GlnIleThrAlaSerSerValLeuGluIleThrAspHisThrGlyGlnGluAsnSerTrp	204
Db	1282	CAATAACAGCATCATCTGTGCTGGAGTGGACTGACCACACAGGGCCAGAGACAGTGG	1341
Qy	205	LysProLysLysAlaArgLeuLysLysProGlyProProTrpPAlaAlaPheAlaThrAsp	224
Db	1342	AAACCCAAAAGCCAGGCTGAATAAACTGACCGCCCTTGGCTGCTTTGGCTGCTTTGGCACTGAT	1401

Qy	225	GlutThrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThr	244
Db	1402	GAATACACAGTGGTTACAAATAGATTGAAATAGGAAAAAGAAAAATAACAGGCAATTATAACC	1461
Qy	245	ThrGlySerThrMetValGluHisAsnTyrTrpValSerAlaTyrArgIleLeuTyrSer	264
Db	1462	ACTGGATCCACCAATGGTGGAGCACAAATACTATGTGCTGCTGCCTACAGAACTCTGTACAGT	1521
Qy	265	AspAspGlyGlnLysThrPrpThrValTyrArgGluProGlyValGluAspLysIlePhe	284
Db	1522	GATGATGGCGAGAAATGGACTGTGTACAGAGCCGTGCTGGGAGCAAGATAAGATATTT	1581
Qy	285	GlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProProlIleAla	304
Db	1582	CAAGGAAACAAGATATCACCAGGATGGCGTAAATAACTTTTTCGCCACCAATTTATTGGCA	1641
Qy	305	ArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeu	324
Db	1642	CGTTTATTAGAGTGAACTCTACCCAAATGGCAGCAAAAATTTGCCATGAAAATGGACCTG	1701
Qy	325	LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProProPro	344
Db	1702	CTCGGATGTCAAGTTTATCTTAAAGGTGCTGCTCCAAAACTTACTCAACCTCCACCTCCT	1761
Qy	345	ArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla	364
Db	1762	CGGAACAGCAANTGACCTCAAAAACACTACAGCCCTCCAAAAATAGCCAAAGTCGTGCC	1821
Qy	365	ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu	384
Db	1822	CCAAAATTTACGCAACCACTACAACTCCGCACTAGCAATGAATTTCTTGCACACAGAA	1881
Qy	385	GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp	404
Db	1882	CAACAACCTGCCAGTCTGTATCATCAGAAATACTACCGTAACCTCCAATAATGAACCAAGAT	1941
Qy	405	ValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu	424
Db	1942	GTAGCGCTGGCTGCAGCTTCTTGTCCTGTGCTGGCTCATGGTCTCCTCCTCCTCATCTC	2001
Qy	425	IleLeuValCysAlaIrrpHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrAsp	444
Db	2002	ATATTAGTGTGCTTGGCACTGGAGAAACAGAAAGAAAAAATCTGAAGGCACCTTATGAC	2061
Qy	445	LeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLys	464
Db	2062	TTACCTTACTGGGCCCGGCGAGTTTGGTGGAAAGGAATGAAGCAGTCTTCTCTCCGAAAA	2121
Qy	465	AlaValAspHisGluGluThrProValArgTyrSerSerSerGluValAsnHisLeuSer	484
Db	2122	GCAGTGGACCATGAGGAAACCCCGATTCGCTATACGACGCGAAGTTTAATCACCCTGAGT	2181
Qy	485	ProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuVal	504
Db	2182	CCAAAGAAAGTCAACACAGTGTGCAGGCTGACTCTGCAGAGATATGCTCAGCCACTGGTA	2241
Qy	505	GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGlu	524
Db	2242	GGAGGAATTTGTGTACACTTCATCAAGATCTACCTTTAAACCAGAAAGGAGGAAAAAGAA	2301
Qy	525	AlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyr	544
Db	2302	GCAGGCTATGCAGACCTTAGATCTTACAACCTCACCAGGCGAGAAAGTTTATCATGCCTAT	2361
Qy	545	AlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetAspMetSer	564
Db	2362	GCTGAACCACTCCCAATACGGGGCTGAGTATGCAACCCCAATCATCATGGACATGTCA	2421
Qy	565	GlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsn	584
Db	2422	GGGCACCCCACTTCAGTTGGTGAAGCCCTCCACATCCACTTTCAAGCCTACGGGGAAC	2481
Qy	585	GlnProProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSer	604

KTGERIRIKFGDFDIEDSDYCHNLNLYKIFNGIGVSRTEIGKYGGLGLOMNOSTESKGS
 EITVLEMSGHIASRGFLASYSVIDKQDLITCLDVTYSNELEPEFSKCPAGCLLPAAE
 ISGTIPHGKDSPLCMAGIHAGVSDVGGQISVVISLKGTPYIESSKLANNVTSWGY
 LSTSLFTFTSGCYGLGMEGVADPQITASSVLSKTDHMGQENSKPEKARLRKPG
 PPWAAFAADKIFQGNKDYHKDVRNRPPIIARFIRVNPVQWOKIAMVELLGCQFTL
 REFGAAQDKIFQGNKDYHKDVRNRPPIIARFIRVNPVQWOKIAMVELLGCQFTL
 KGRLPKLTOPPPRNNLNKNTTVHPKLGRAKFTOALOPRSRNDLPLLPAAOTATPD
 VKNTTVPSTVKDVALAANVLPVLYVALTLLILLYCALHWRNRKKKASGTYDLPHWD
 RAGWGVKDLPAKSPVEHEETPVRSNVEVLSHSPREVTIVLQADSAEQAQLVGGI
 VGTLHORSTFKPEBGEKASVADIDPNAPVQEVYHAYAEPLPVTGPEYATPIVMDMSG
 HSTASVGLPSTFTFRTAGNQPALVGTNTLLSRDSCSSGQAQYDTPKGGKPAAPAE
 ELVTVQPSQTOEASGAGRDEKDFAKETIL"

BASE COUNT 612 a 606 c 594 g 527 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.97e-195 Length: 2339
 Score: 2984.50 Matches: 566
 Percent Similarity: 83.26% Conservative: 36
 Best Local Similarity: 78.28% Mismatches: 42
 Query Match: 86.01% Indels: 79
 DB: 10 Gaps: 3

us-10-060-830-1114 (1-653) x AF387549 (1-2339)

Qy 7 LeuLeuValLeuLeuLeuLeuGluAspAlaGlyAlaGlnGln----- 22
 Db 159 CTGCTCTTCTACTGCTCTACTGCTCCGGACGCTGGAGCCCAAGAGGTGATGGATG 218
 Qy 22 ----- 22
 Db 219 GGACACACTGTACTAGGCCCTGACAGTGGAGAACCTTACATCCATCACTCCACATACC 278
 Qy 22 ----- 22
 Db 279 TATCTAACACTACTGTGTGTAATGGAGATTCCAGTAAAGACGGGAGAAAGAAATCCGC 338
 Qy 22 ----- 22
 Db 339 ATCAAGTTCGGTGACTTTTGACATTGAAGATTCTGATTATTGTCACCTTAATTAACCTGAAA 398
 Qy 23 -----GlyLysTyrCysGlyLeuGlyLeu 30
 Db 399 ATCTTTAATGAANTGGACTCAGCAGACGGAATAATAGCAAGTACTGTGCTGGGTTA 458
 Qy 31 GlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGly 50
 Db 459 CAATGAATCAGTCAATTTAGTCCAAAGGCGAGTGAATCAGAGTGTCTCATGAGTGA 518
 Qy 51 IleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeu 70
 Db 519 ATCCATGCTTCTGGTGGAGGATTTTGGCTTCTTACTCAGTTATAGATAACAAGATTTA 578
 Qy 71 IleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCysPro 90
 Db 579 ATCACTTGTGGATAGTATATCTAATTTTGGAACTGAGTTCAGTAACTAGTCCCA 638
 Qy 91 AlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAsp 110
 Db 639 GCTGGCTGCTGCTGCTCTTGGCTGGAATATCTGGAACGATTCTCATGGATATAGAT 698
 Qy 111 SerSerProLeuCysMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGly 130
 Db 699 TCTTACCGCTGTGTATGGCTGGAAATCCATCAGGAGTAGTGTGATGATGCTGGGTGGC 758
 Qy 131 GlnIleSerValIleSerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnAsn 150
 Db 759 CAATATCCGCTTGTGATTAAGAGGACCCCATATATACGAAGATTCTTTGGCCCAACAT 818
 Qy 151 ValThrSerValGlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCys 170
 Db 819 GTCACCTCCATGGTGGGATCTTATCTACGAGTCTGTTTACATTTAAGACAAAGTGGTTGC 878

Qy 171 TyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSer 190
 Db 879 TATGGGACTCTAGGGATGGAGTTCAGGTGTGATCCCGGATCCCCAGATTAACAGCATCATCT 938
 Qy 191 ValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArg 210
 Db 939 GTACTGGAGTGGAGTACCACATCGGGCAGGAGAACAGCTGGAACCCGAGAGGCCAGG 998
 Qy 211 LeuLysLysProGlyProProTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGln 230
 Db 999 CTGAGAAAACGGGGCCCTCCCTGGGCTGCTTTGGCACTGATGAGCATCACTGGCTGCA 1058
 Qy 231 IleAspLeuAsnLysGluLysLysIleThrGlyIleThrThrGlySerThrMetVal 250
 Db 1059 ATTGACCTTAATAGGAGAGAAAGATACAGGCATCGTAACCACTGGATCTACCCCTGATA 1118
 Qy 251 GluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspAspGlyGlnLysTrp 270
 Db 1119 GAGCACAAATTACTATGTCTGCTACAGAGTTCTGTACACTGACGATGGCAGAAATGG 1178
 Qy 271 ThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyr 290
 Db 1179 ACTGTGTACAGAGAGCCTGCTGGCTCAGCAGCAAGATATTTCAAGGAAACAAAGATTAT 1238
 Qy 291 HisGlnAspValArgAsnAsnPheLeuProProIleAlaArgPheIleArgValAsn 310
 Db 1239 CACAAAGATGTCTGATAAATCTTTTGGCCCAATATTGACCGTTTCATTAGAGTGAAC 1298
 Qy 311 ProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheIle 330
 Db 1299 CCTGCCAGTGGCAACAGAAAATGGCCATGAAGTGAATGCTGGGATGTCACTTCACT 1358
 Qy 331 ProLysGlyArgProProLysLeuThrGlnProProProArgAsnSerAsnAspLeu 350
 Db 1359 CTGAAGGTCCCTTCCAAAGCTTACTCAACTCCCGCCACCTCGGAGAACAAATACCTC 1418
 Qy 351 LysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPheThrGlnPro 370
 Db 1419 AAAAACTACAGTTCATCCCAACTA-----GTCGTGCCCTCTAAATTTACTCAGGA 1472
 Qy 371 LeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerPro 390
 Db 1473 CTCCAACCCAGCAAGTAGGAATGACCTTCTCTGCTGCCGCCGACAGCAACTGCCACTCT 1532
 Qy 391 AspIleArgAsnThrThrValThrProAsnValThrLysAspValAlaLeuAlaVal 410
 Db 1533 GATGTCAAAAACACGACTGTGACTCCCAAGTGTGACCAAAAGATGTTGCACCTGGCCCGCT 1592
 Qy 411 LeuValProValLeuValMetValLeuThrThrLeuIleLeuLeuValCysAlaTrp 430
 Db 1593 CTGTTCTCTGTGCTGGTCAATGGCCCTCACACACTCATCTCTATCTAGTGTGCTGG 1652
 Qy 431 HisTrpArgAsnArgLysLysLysThrGlyThrTyrAspLeuProTyrTrpAspArg 450
 Db 1653 CATTTGGAGAACAGAAAGAAAAAGCCGAGGCACTTATGATTTACCCCACTGGGATCGG 1712
 Qy 451 AlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGlu 470
 Db 1713 GCAGGCTGTGGAAAGGAGTAGTGTCTCCCTGTCCTGCAATTCCTGGTGAACACAGGAG 1772
 Qy 471 ThrProValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrThr 490
 Db 1773 ACGCCAGTGCCTACAGCAACAGTGAAGTTAGTACCTTACGCCCCGAGGAGGAGTCAAGCA 1832
 Qy 491 ValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThr 510
 Db 1833 GTGCTGAAGCTGATCTGCGAGATACCCAGACCCCTCGTGGGAGGAATTTGTCGACACA 1892
 Qy 511 LeuHisGlnArgSerThrPheLysProGluGlyLysGluAlaGlyTyrThrAlaAspLeu 530
 Db 1893 CTCCATCAGAGATCTACCTTTAAACCTGAAGAAGGAAAAAGAGCGAGCTACGAGATCTA 1952
 Qy 531 AspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGlyProLeuProfile 550

Db	1953	GACCCCTACAAACGCTCCAGTACAGGAAGTGTATCATGCCCTACGCTGAGCCCTGCCGGTA	2012
QY	551	ThrGlyProGluTyrAlaThrProIlelleMetAspMetSerGlyHisProThrThrSer	570
Db	2013	ACGGGCGCTGAGTACGCAACCCCAATCGTCATGAGCATGTTCAGGCGCATCCACAGCCTCA	2072
QY	571	ValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProProProLeuVal	590
Db	2073	GTTGGTGTGCGCCCTCCACATCCACTTTCAGAACTGTCAGGGAACCAAGCTCCCGCATTAATG	2132
QY	591	GlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyr	610
Db	2133	GGAACTTACAACTCTCTCTCCAGAGCTGACAGCTGTTCCTCGGCGCAGGCTCAGTAC	2192
QY	611	AspThrProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTyrGlnVal	630
Db	2193	GACACCCCAAAAGTGGGAGCCAGCA--GCTGCCCCACAGGGAACCTGGGTACCAAGTG	2249
QY	631	ProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGlyGluCysAspValPheLys	650
Db	2250	CCGAGAGCAGCCAGGAAGCATCAGGAGCAGGAAGGATGAGAAATTTGATGCTTTTAAA	2309
QY	651	GluIleLeu	653
Db	2310	GAAACCCCTT	2318
RESULT 5	AF387548		
LOCUS	AF387548	2642 bp mRNA linear	ROD 11-NOV-2001
DEFINITION	Mus musculus strain ICR endothelial and smooth muscle cell-derived		
ACCESSION	AF387548	neuropilin-like protein (Esdn) mRNA, complete cds.	
VERSION	AF387548.1	GI:16902436	
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 2642)		
TITLE	Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N., Matsumori, A., Sasayama, S., Honjo, T. and Tashiro, K.		
JOURNAL	ESDN, a novel neuropilin-like membrane protein cloned from vascular cells with the longest secretory signal sequence among eukaryotes, is up-regulated after vascular injury		
MEDLINE	J. Biol. Chem. 276 (36), 34105-34114 (2001)		
PUBMED	21424027		
AUTHORS	2 (bases 1 to 2642)		
TITLE	Kobuke, K., Furukawa, Y., Sugai, M., Tanigaki, K., Ohashi, N., Sasayama, S., Matsumori, A., Honjo, T. and Tashiro, K.		
JOURNAL	Direct Submission		
FEATURES	Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan		
source	Location/Qualifiers		
gene	1. .2642		
CDS	/organism="Mus musculus"		
	/strain="ICR"		
	/db_xref="taxon:10090"		
	1. .2642		
	/gene="Esdn"		
	/notes="similar to the Mus musculus strain 129/SvJ Esdn gene sequence in GenBank Accession Number AF442820"		
	314. .2623		
	/gene="Esdn"		
	/codon_start=1		
	/product="endothelial and smooth muscle cell-derived neuropilin-like protein"		
	/protein_id="AA130179.1"		
	/db_xref="GI:16902437"		
	/translation="MASRAPFLRAARSPQGGPAAAPAAATGRAALPSAGCCPLPGRNFS		
	SSRPRLILLILLIQAGQGGCGTGLGPGSGTLTSINPTYPNSTVCEWEIWEI		

Qy	151	ValThrSerValValGly#HisLeuSerThrSerLeuPheThrPheLysThrSerGlyCys	170
Db	1121	GTCACTTCCACGCTGGGATACTTATCTCAAGTCTGTTTACATTTAAGACAAGTGTGTC	1180
Qy	171	TyrGlyThrLeuGlyMetGlnSerGlyValIleAlaAspProGlnIleThrAlaSerSer	190
Db	1181	TATGGAGCTCTCGGATGGAGTCTGTGTGATGCCCATCCAGATAACAGANTCGTCT	1240
Qy	191	ValLeuGluTrpThrAspHisThrGlyGlnGlnAsnSerTrpLysProLysLysAlaArg	210
Db	1241	GCACCTGGAGTGCACCATGGGCGCAGGAGAACGCTGGACAGCGGAGAGGCCAGS	1300
Qy	211	LeuLysLysProGlyProProTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGln	230
Db	1301	CTGAGAAACCCGGGCTCCCTGGGCTGCTTTTGCCACTGATGACATAGTGCCTGCAG	1360
Qy	231	IleAspLeuAsnLysGluLysLysIleThrGlyIleIleThrThrGlySerThrMetVal	250
Db	1361	ATAGACCTTTAAACAAGGAGAAGAATACAGTATCTGTAACCACTGGGTCTACCATGATA	1420
Qy	251	GluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspAspGlyGlnLysTrp	270
Db	1421	GAACACAGTTACTATGTCTGCTCCCTACAGAGTCTGTACAGTACGATGGCAGAGATGG	1480
Qy	271	ThrValTyrArgGluProGlyValGlnAspLysIlePheGlnGlnLysAsnLysAspTyr	290
Db	1481	ACTGTGTACAGAGAACTCTGGTGTGGCAGGACAGAATATTTCGAAGAAACAAGATTAT	1540
Qy	291	HisGlnAspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgValAsn	310
Db	1541	CACAAGGATGTCGTAATAACTTTTGGCCACCAATTATGTCACGTTTCATTAGAGTGAAC	1600
Qy	311	ProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheIle	330
Db	1601	CCTGCTCCAGTGGCAACAGAAATTCGCTGAAGTGAAGTGCCTCGGATGTCACTTTACT	1660
Qy	331	ProLysGlyArgProProLysLeuThrGlnProProProArgAsnSerAsnAspLeu	350
Db	1661	CTCAAAAGGTGCCTCCAAAGCTTACT-----CCACCTCTCGGAAGCGCAATAACCTC	1714
Qy	351	LysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPheThrGlnPro	370
Db	1715	AGAAATACTACAGCTCGTCCCAACTAGTGAAGTCTGGCCCTAAATTTACTCAAGTG	1774
Qy	371	LeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerPro	390
Db	1775	CTCCAACTCGAAGTAGGAATGAACCTCTGTGTCAGCGCGGAGACAACTACCACCTCT	1834
Qy	391	AspIleArgAsnThrThrValThrProAsnValThrLysAspValAlaLeuAlaAlaVal	410
Db	1835	GATATAAAAACACGACTCTGACTCCAGCTGAACCAAGATGTGCGACTGGCTGCCGTT	1894
Qy	411	LeuValProValLeuValMetValLeuThrThrLeuIleLeuValLysAlaTrp	430
Db	1895	CTGCTCCTCTGTGTGTCATGGCCCTCACCACTATCTCTCTCTCTCTCTCTCTCTCT	1954
Qy	431	HisTrpArgAsnArgLysLysLysThrGluGlyThrThrThrAspLeuProTyrTrpAspArg	450
Db	1955	CACCTGGAGAAACAGGAAGAAGAAACTGAAGCGCTATGATTTACCCCACTGGGATCGG	2014
Qy	451	AlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGlu	470
Db	2015	GCAGTTGGTGGAAAGAAATGAACAGAGCTTCTCCCTGCCAAGTCTGGTGGACGAGAG	2074
Qy	471	ThrProValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrThr	490
Db	2075	ACGCCAGTGGCTATACACACTAGTGAAGTCAGTCACTGTAGTGGCCAGGAAGTCAACACA	2134
Qy	491	ValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThr	510
Db	2135	GTGCTGCAGCGCCGACTCTGCAGAAATATGCACGCCCTCTGCTGGGAGAAATGTGTGGCACA	2194
Qy	511	LeuHisGlnArgSerThrPheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeu	530

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 21, 2003, 01:52:19 ; Search time 65.278 seconds
(without alignments)
9217.500 Million cell updates/sec

Title: US-10-060-830-1113
Perfect score: 1962
Sequence: 1 atgcctctgtctctctgct.....tttttaagaagaatcctttga 1962

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCRUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	544.6	27.8	640	4	US-09-328-111-480
2	91.6	4.7	6585	3	US-08-746-111-4
3	86.6	4.4	6909	2	US-08-804-196-1
4	86.6	4.4	6909	2	US-08-658-340-1
5	86.6	4.4	6909	3	US-08-746-111-26
6	86.4	4.4	2303	2	US-08-480-229C-9
7	86.4	4.4	2303	2	US-08-659-235C-9
8	78.6	4.0	318	2	US-08-480-229C-19
9	78.6	4.0	318	2	US-08-659-235C-19
10	78.2	4.0	1780	2	US-08-480-229C-11
11	78.2	4.0	1780	2	US-08-659-235C-11
12	76.8	3.9	1384	1	US-07-607-538C-1
13	76.8	3.9	1384	2	US-08-162-402B-1
14	76.8	3.9	1934	2	US-08-162-402B-7
15	63.4	3.2	4629	2	US-08-484-891-7
16	63.4	3.2	4670	3	US-08-717-294-41
17	63.4	3.2	4999	4	US-09-470-618-14
18	63.4	3.2	4999	4	US-09-364-862-14
19	63.4	3.2	5035	2	US-08-882-083-1
20	63.4	3.2	5035	2	US-08-558-107-1
21	63.4	3.2	5035	3	US-09-243-539-1
22	63.4	3.2	6999	1	US-08-276-594A-1
23	63.4	3.2	7056	1	US-08-121-202-1
24	63.4	3.2	7493	1	US-08-212-133A-7
25	63.4	3.2	7493	1	US-08-474-503-5
26	63.4	3.2	7493	2	US-08-670-707A-5
27	63.4	3.2	7493	4	US-09-037-601-5

28 63.4 3.2 7493 4 US-09-315-179-5 Sequence 5, Appli
29 63.4 3.2 7493 5 PCT-US94-13200-5 Sequence 5, Appli
30 63.4 3.2 8241 6 5171844-1 Patent No. 5171844
31 63.4 3.2 8967 1 US-08-366-851A-1 Sequence 1, Appli
32 63.4 3.2 9009 1 US-07-864-004B-3 Sequence 3, Appli
33 63.4 3.2 9009 1 US-08-251-937A-3 Sequence 3, Appli
34 63.4 3.2 9009 1 US-08-212-133A-1 Sequence 1, Appli
35 63.4 3.2 9009 1 US-08-474-503-1 Sequence 1, Appli
36 63.4 3.2 9009 2 US-08-670-707A-1 Sequence 1, Appli
37 63.4 3.2 9009 4 US-09-037-601-1 Sequence 1, Appli
38 63.4 3.2 9009 4 US-09-315-179-1 Sequence 1, Appli
39 63.4 3.2 9009 4 US-09-523-656-1 Sequence 1, Appli
40 63.4 3.2 9009 5 PCT-US93-03275-3 Sequence 3, Appli
41 63.4 3.2 9009 5 PCT-US94-13200-1 Sequence 1, Appli
42 63.4 3.2 9354 1 US-08-683-839B-2 Sequence 2, Appli
43 63.4 3.2 11933 4 US-09-470-618-13 Sequence 13, Appli
44 63.4 3.2 11933 4 US-09-364-862-13 Sequence 13, Appli
45 63.2 3.2 7032 4 US-09-324-867-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-328-111-480/c
; Sequence 480, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-480

Query Match 27.8%; Score 544.6; DB 4; Length 640;
Best Local Similarity 96.7%; Pred. No. 5.2e-167;
Matches 616; Conservative 0; Mismatches 15; Indels 6; Gaps 6;
Qy 893 TTTTGGCCACCAATATTTCACGCTTTTATAGAGTAACTACCAATGCGACGAGAAA 952
Db 632 TTTTGGCCACCAATATTTCACGCTTTTATAGAGTAACTACCAATGCGACGAGAAA 576
Qy 953 TTGGCCATGAAATGAGCTGCTCGGATGTCAGTTTATTCCTAAAGTCTCTCTCCAAAAC 1012
Db 575 TTGGCCATGAAATGAGCTGCTCGGATGTCAGTTTATTCCTAAAGTCTCTCTCCAAAAC 518
Qy 1013 TTACTCAACCTCCACCTCTCGGAACAGCAATGACCTAAAAACACTACAGCCCTCCAA 1072

Db 517 TTAACAACCTCCACCTCTTNGAAGCAAGCAATGNCCTCAAAACCACTACAGCCCTCC-A 459
QY 1073 AATAGCCAAAGTGTGTCGCCCAAAATTTACGCAACCACTACAACTTCGAGTAGCAATG 1132
Db 458 AATAGCCAAAGTGTGTCGCCCAAAATTTACGCAACCACTACAACTTCGAGTAGCAATG 399
QY 1133 AATTTCTCTGACAGACAGAAACAACTGCCAGTCTGTATCAGAAATACCTACCGTAA 1192
Db 398 AATTTCTCTGACAGACAGAAACAACTGCCAGTCTGTATCAGAAATACCTACCGTAA 339
QY 1193 CTCAATGTAAACAAAGATGTAGCGTGGCTGAGTTCCTTCCCTGCTGCTGCTATGG 1252
Db 338 CTCAATGTAAACAAAGATGTAGCGTGGCTGAGTTCCTTCCCTGCTGCTGCTATGG 279
QY 1253 TCCTCACTACTCTCATTTCTATATTAGTGTGCTTGGCACTGGAGAAACAGAAAGAAA 1312
Db 278 TCCTCACTACTCTCATTTCTATATTAGTGTGCTTGGCACTGGAGAAACAGAAAGAAA 219
QY 1313 AAACGTGAAGCACCCTATGACTTACTTACTTGGGACCGGGCAGGTTGGTGGAAAGGAATGA 1372
Db 218 AAACGTGAAGCACCCTATGACTTACTTACTTGGGACCGGGCAGGTTGGTGGAAAGGAATGA 159
QY 1373 AGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAAACCCAGTTCGCTATAGCAGCA 1432
Db 158 AGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAAACCCAGTTCGCTATAGCAGCA 99
QY 1433 GCGAAGTAAATCACTGAGTCCCAAGAGAGTCAACAGTCTGCTGAGGCTGACTCTGCAG 1492
Db 98 GCGAAGTAAATCACTGAGTCCCAAGAGAGTCAACAGTCTGCTGAGGCTGACTCTGCAG 39
QY 1493 AGTATGCTCAGCCACTGTAGGAGGAATTTGTTGGTAC 1529
Db 38 AGTATGCTCAGCCACTGTAGGAGGAATTTGTTGGTAC 2

RESULT 2

US-08-746-111-4
; Sequence 4, Application US/08746111
; Patent No. 6066778
; GENERAL INFORMATION:
; APPLICANT: Ginsburg, David
; APPLICANT: Cui, Jisong
; TITLE OF INVENTION: Compositions And Methods For Screening
; TITLE OF INVENTION: Compounds For Anticoagulant Activity
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746.111
; FILING DATE: 06-NOV-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6585 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6...6554
; US-08-746-111-4
Query Match 4.7%; Score 91.6; DB 3; Length 6585;
Best Local Similarity 53.3%; Pred. No. 7e-19;
Matches 193; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 650 CTGGGGCTGCTTTTGGCCACTGATGATACCAAGTGGTGGAGCACAATTTCTCAATAGGAAA 709
Db 6214 CTGGGCAAGCCAGGCAAAACAAACAAAGCAGTGGTTTACAGTCGATCTGCTCAAAATCA 6273
QY 710 AGAAATAAAGCAGGCAATTAACCACTGCATCCACATGGTGGAGCACAATTTCTATGTGT 769
Db 6274 AGAAGGTAAAGCGGCATCGTAACGCGAGGCTGTAGTCTCTGCTCTGAGATGTACGTGA 6333
QY 770 CTGCTACAGAAATCCTGTACAGTGTATGGCAGAAATGAGTGTGTACAGAGCGCTG 829
Db 6334 AGACTACAGCATCCAGTACAGTACCCAGGCTGGCATGGAACCTTACCCAGAGAAAT 6393
QY 830 GTGTGGAGCAAGATAAGATATTCAAGAAACAAAGATTTATCACCAGATGTGCTTAATA 889
Db 6394 CCTCCATGGTGGCAAGATTTTGAAGGAAACAGCAATACCAAGGGGCACATGAAGAACT 6453
QY 890 ACTTTTGGCCACCAATATTTCAGCTGTTTATTAGAGTGAATCTTACCAATGGCAGCAGA 949
Db 6454 TTTTCAACCCGCCCATTAATTTCCAGATTTATCCGATCATTCTTAAACATGGAAACCAGA 6513
QY 950 AAATTTGCCATGAAATGAGCTGCTCGGATGCTCAGTTTATTCTTAAAGGTGCTCCTCCAA 1009
Db 6514 CATCGCCCTTCGCTAGAGCTCTTCGCTGTGACATTTATTAGAAATTAATTTCCAAAAA 6573
QY 1010 AA 1011
Db 6574 AA 6575

RESULT 3

US-08-804-196-1
; Sequence 1, Application US/08804196
; Patent No. 5874256
; GENERAL INFORMATION:
; APPLICANT: Bertina, Rogier
; APPLICANT: Reitsma, Pieter
; TITLE OF INVENTION: A method for diagnosing an increased
; TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
; TITLE OF INVENTION: thrombosis and kit for use with the same.
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5874256el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.196
; FILING DATE:
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/454.353
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409

TITLE OF INVENTION: Compositions And Methods For Screening
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESS: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,111
FILING DATE: 06-NOV-1996

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02536
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 6909 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna

US-08-746-111-26

Query Match 4.4%; Score 86.6; DB 3; Length 6909;
Best Local Similarity 51.7%; Pred. No. 3.1e-17;
Matches 259; Conservative 0; Mismatches 224; Indels 18; Gaps 2;

Qy 493 TTTAAGCAAGTGGATGTTATGGAACACTGGGATGGGATGGTGTGTCATCGCGATCCT 552

Db 6271 TGTAGGTAATGGATGTTCCACACCCCTGGTATGGAATGGAAGATAGAAAACAG 6330

Qy 553 CAATAACAGCATCATCTGCTGGAGTGGTACACACAGGCGCAAGCAAGCAAGTTGG 612

Db 6331 CAATCACAGCTCTTC-----GTTAAGAATCTTGGTGGGAGATTACTGG 6378

Qy 613 AAACCCAAAAAGCCAGGCTGAAAAACCTGGAC-----CGCCCTGGGCTGCTTTTCC 666

Db 6379 GRACCTTCCGTCGCGCTGAATGCCAGGACGCTGTGAATGCCCTGGCAAGCAAGCA 6438

Qy 667 ACTGATGAATACCACTGGTTCACAAATAGATTGTAAGGAAAGAAATACAGGCATT 726

Db 6439 AACAAATAAGCAGTGGCTAGAAATTTGATCTACTCAAGATCARGAAGATAACGGCAAT 6498

Qy 727 ATAACCACTGGATCCACCATGGTGGAGCACAATTAATGTCCTGCTACAGAAATCCTG 786

Db 6499 ATAACACAGGGCTGCAAGTCTCTGCTCTGAAATGTAAGAGCTATACCATCCAC 6558

Qy 787 TACAGTGATGATGGGAGAAATGGACTGTGTACAGAGACCCCTGGTGTGGAGCAAGATAG 846

Db 6559 TACAGTGAGCAGGAGTGGATGGAATGGAACCATACAGCTGAAATCTCCATGTGGACAAG 6618

Qy 847 ATATTTCAAGGAACAAAGATTATACAGGATGTCGCTAAATACCTTTTGGCCCAAT 906

Db 6619 ATTTTGAAGGAATACTAATACCAAGGACATGTGAAAGAACTTTTCAACCCCCCAAT 6678

Qy 907 ATTGCAGGTTTATAGAGTGAATCTTACCAATGCGAGCAAGAAATGCGATGAAATG 966

Db 6679 ATTTCCAGGTTTATCCGTGTCATTTCTTAACAAATGGAATCAAGATATACACTTCG 6738

Qy 967 GAGCTGCTCGGATGTCAGTTT 987

Db 6739 GAACTCTTTGGCTGTGATATT 6759

RESULT 6

US-08-480-229C-9
Sequence 9, Application US/08480229C
Patent No. 5874562

GENERAL INFORMATION:
APPLICANT: Quertermous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.

TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,229C
FILING DATE: 07-JUN-1995

CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999

TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cdna
FEATURE:

NAME/KEY: CDS
LOCATION: 619..2058
US-08-480-229C-9

Query Match 4.4%; Score 86.4; DB 2; Length 2303;
Best Local Similarity 50.2%; Pred. No. 1.7e-17;
Matches 241; Conservative 0; Mismatches 236; Indels 3; Gaps 1;

Qy 507 ATGTATGGAACACTGGGATGGATCTCTGTGATCGGGATCCTCAATAACAGCATC 566

Db 1089 ATGCTCTGGGCATTGGGAATCGAGGTGGATCATATCATATCAGCAATCAGCTTC 1148

Qy 567 ATCTGTGCTGGAGTGGACTGACCCACACAGGCAAGAGACAGTTGGAACCCCAAGC 626

Db 1149 ATCTATATC---ACCGAGCTCTTTTGGACTCCAGAAAGTGTATCCCTACTATCTCGACT 1205

Qy 627 CAGGCTGAAAAACCTGGAGCCGCTTTGGGCTGCTTTTGGCACTGATGATACCACTGGTT 686

Db 1206 TAATAAGAGGGCCTTATAAATGCTGGACAGCTGTGAAATGACAGATGGCCATGGAT 1265

Qy 687 ACAATAGATTTCAATAAGGAAAGAAATAACAGCAATTAACCACTGATGATACCACTCCACAT 746

Db 1266 TCAGATAAATTTGCAAGAAAAATGAGATCTCTGGTGTATTACCCAGGCAAGAAAG 1325

Qy 747 GGTGGAGCAATTAATCTATGTCTGCTACAGAAATCCTACAGTGTATGGGCAAG 806

	Query Match	4.0%;	Score 78.6;	DB 2;	Length 318;
	Best Local Similarity	53.0%;	Pred. No. 1.5e-15;		
	Matches 168;	Conservative	0;	Mismatches 149;	Indels 0;
QY	670	GATGAAATACCAGTGGTTTACAAATAGATTTGAATTAAGGAAAGAAAATAACAGGCATTATA	729		
DB	1	GACAGATGGCCATGGATTCAGATTAATTTGCAAGAAAAATGAGAGTCACATGGTGTATT	60		
QY	730	ACCAGTGGATCCACCATGGTGGAGCACAAATTAATTAATGTGTCTGCCTCAGAAATCCCTGTAC	789		
DB	61	ACCAAGGAGCAAAAAGGATTGGAAGCCAGAGTACATAAAATCCCTACAAAATTCGCCATC	120		
QY	790	AGTGATGATGGCGAGAAATGCACTGTGTACAGAGAGCCGTGTGTGGAGCAAGATAAGATA	849		
DB	121	AGCAATGACGGGAAGACCTCGGGCAATGTACAAAGTAAAGGCCACCAATCAAGAGATGGTC	180		
QY	850	TTTCAAGAAACAAGATTTATCACCAGGATGTGCCGTAATAACTTTTTGGCCACCAATTAAT	909		
DB	181	TTTCGTGGAAATGTGTGATTAACAACACACCATATGCTAATCTTTCACACCCCCCAATCAAA	240		
QY	910	GCAGCTTTTATTAGAGTGAATCCTTACCNAATGGCAGCAGAAAATTTGCCATGAATGGAG	969		
DB	241	GCTCAGTATGTAAGACTTACCCCCAAAATTTGCGAAGGCAATTTGTACTTTAAGAAATGSA	300		
QY	970	CTGCTCGGATGTCACTT	986		
b	301	CTTCCTGGCTGTGAGCT	317		

RESULT 10
US-08-480-229C-11
Sequence 11, Application US/08480229C
Patent No. 5874562
GENERAL INFORMATION:
APPLICANT: Quertermous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/480,229C
APPLICATION NUMBER: 07-1995
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:

RESULT 10
US-08-480-229C-11
Sequence 11, Application US/08480229C
Patent No. 5874562
GENERAL INFORMATION:
APPLICANT: Quertermous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/480,229C
APPLICATION NUMBER: 07-1995
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:

REFERENCE/DOCKET NUMBER: CRFCC-004

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 943-1931
; TELEFAX: (510) 943-1189
; TELEX: N.A.
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1384 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-07-607-538C-1

Query Match          3.9%; Score 76.8; DB 1; Length 1384;
Best Local Similarity 51.5%; Pred. No. 1.6e-14;
Matches 177; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 650 CTTGGGCTGCTTTTGGCCACTGATGAATACACAGTGGTTACAAATAGATTGTAATAGGAAA 709
DB 323 CTTGGGCTGCTTTTGGGAGCTACGGTAACGATCAGTGGCTGCGAGTGGACCTCGGCTCTCGA 382
QY 710 AGAAATACAGGCATATTAACACATGATGGCAGAAATGATGATGATGATGATGATGATGATG 769
DB 383 AGAGGTGACAGGCATCATACCCAGGGGCGCGTAACCTTGGCTCTGCTCCAGTTGTGG 442
QY 770 CTGCTACAGAACTCTGACAGTGAATGATGGCAGAAATGATGATGATGATGATGATGATGATG 829
DB 443 CATCTTACAGAGTGGCTACAGTAATGACAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGT 502
QY 830 GTGTGGAGCAAGATGAATATTTCAAGGAAACAAAGATTAATACCAATGATGATGATGATGATG 889
DB 503 GGACTGGCAGCAGTATGATCTTCCCTGGCAACTGGGAGCAACCACTCCCAAGAAGAACT 562
QY 890 ACTTTTGGCCACCAATATTGACAGTGGTTCGATGATGATGATGATGATGATGATGATGATG 949
DB 563 TGTGTGAGAGCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 622
QY 950 AAATTGGCCATGAATGAGTGGTTCGATGATGATGATGATGATGATGATGATGATGATGATG 993
DB 623 GCATCGCCCTGCGCTGGAGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCT 666

RESULT 13
US-08-162-402B-1
; Sequence 1, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; GLOBULE (HMF) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 435
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1384 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-162-402B-1

Query Match          3.9%; Score 76.8; DB 2; Length 1384;
Best Local Similarity 51.5%; Pred. No. 1.6e-14;
Matches 177; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 650 CTTGGGCTGCTTTTGGCCACTGATGAATACACAGTGGTTACAAATAGATTGTAATAGGAAA 709
DB 323 CTTGGGCTGCTTTTGGGAGCTACGGTAACGATCAGTGGCTGCGAGTGGACCTCGGCTCTCGA 382
QY 710 AGAAATACAGGCATATTAACACATGATGGCAGAAATGATGATGATGATGATGATGATGATG 769
DB 383 AGAGGTGACAGGCATCATACCCAGGGGCGCGTAACCTTGGCTCTGCTCCAGTTGTGG 442
QY 770 CTGCTACAGAACTCTGACAGTGAATGATGGCAGAAATGATGATGATGATGATGATGATGATG 829
DB 443 CATCTTACAGAGTGGCTACAGTAATGACAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGT 502
QY 830 GTGTGGAGCAAGATGAATATTTCAAGGAAACAAAGATTAATACCAATGATGATGATGATGATG 889
DB 503 GGACTGGCAGCAGTATGATCTTCCCTGGCAACTGGGAGCAACCACTCCCAAGAAGAACT 562
QY 890 ACTTTTGGCCACCAATATTGACAGTGGTTCGATGATGATGATGATGATGATGATGATGATG 949
DB 563 TGTGTGAGAGCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 622
QY 950 AAATTGGCCATGAATGAGTGGTTCGATGATGATGATGATGATGATGATGATGATGATGATG 993
DB 623 GCATCGCCCTGCGCTGGAGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCTGCTGGGCT 666

RESULT 14
US-08-162-402B-7
; Sequence 7, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; GLOBULE (HMF) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 435
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
```

```

; APPLICATION NUMBER: US/08/484,891
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/218,335
; FILING DATE: 25-MAR-1994
; APPLICATION NUMBER: 08/074,920
; FILING DATE: 10-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 271010-273
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4629 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA primer
; FEATURE:
; NAME/KEY: Factor VIII cDNA with
; NAME/KEY: B domain deleted
; US-08-484-891-7

Query Match 3.2%; Score 63.4; DB 2; Length 4629;
Best Local Similarity 51.6%; Pred. No. 8.4e-10;
Matches 173; Conservative 0; Mismatches 156; Indels 6; Gaps 1;

QY 650 CTGTGGCTGCTTTTGGCCACTGATGAATACACAGTGGTTTACAAATAGATTTGAATAAGAAA 709
DB 4070 CTTGGAGACCTCAGGTGAATAATCCRAAGAGTGGCTGCAAGTGACTTCAGAAGACAA 4129
QY 710 AGAAATACAGGCATTATTAACCACTGGATCCACCATGGTGGGAGCACAAATTACTATGTGT 769
DB 4130 TGAAGTTCACAGGAGTAACCTACTCAGGAGATAAAATCTCTGCTTACCAGCATGTATGTGA 4189
QY 770 CTGCCTACAGAAATCCTGTACAGTGTATGCGGAGAAATGGACTGTGTACAGAGCGCTG 829
DB 4190 AGGAGTTCCTCATCTCCAGAGCAAGATGCCATCGACTCTCTTTTTCAGAAATG 4249
QY 830 GTGTGGAGCAAGATAAGATATTTCAAGGAAACAAAGATTATCACACGAGTGTGCGTAAATA 889
DB 4250 GCA-----ANGTAAGGTTTTTTCAGGGAATCAGACTCCTTCACACCTGTGTGACT 4303
QY 890 ACTTTTTTGGCCACCAATATTATTCACGTTTATTATTAGAGTGAATCTTACCCAATGGCAGAGA 949
DB 4304 CTCTAGACCCACCGTTTACTGACTCGCTACCTTCGAATTACCCCCCAGAGTTGGGTGCACC 4363
QY 950 AATTGCCATGAATATGGAGCTGCTCGGATGTCTAG 984
DB 4364 AGATTGCCCTTGAGGATGGAGTTCTTGGGCTGCGAG 4398

```

Search completed: January 21, 2003, 09:28:16
Job time : 125.278 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 09:28:20 ; Search time 68.9384 Seconds
(without alignments)
11728.275 Million cell updates/sec

Title: US-10-060-830-1113

Perfect score: 3559

Sequence: 1 atgcctctgtctctctgct.....ttttaagaatactttgta 1962

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh
-Q/cgn2_1/USPTO_SPOOL/US10060830/runat_16012003_091239_26852/app_query.fasta_1.6933
-DB=SPTRMBL_21 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptn -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn_1_1_357@runat_16012003_091239_26852 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_21:
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3422	96.2	775	4	Q96PD2 homo sapien

ID	Q96PD2	PRELIMINARY	PRT	775 AA.
AC	Q96PD2			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Endothelial and smooth muscle cell-derived neuropilin-like protein.			
GN	ESDN.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	PubMed-11447234;			
RA	Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,			
RA	Matsumori A., Sasayama S., Honjo T., Tashiro K.,			
RT	"ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular			
RT	Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is			
RT	Up-regulated after Vascular Injury."			
RL	J. Biol. Chem. 276:34105-34114(2001).			
CC	-!- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL REGULATORS.			

ALIGNMENTS

RESULT 1

Q96PD2 PRELIMINARY; PRT; 775 AA.

AC Q96PD2
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
GN ESDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RA PubMed-11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.,
RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury."
RL J. Biol. Chem. 276:34105-34114(2001).
CC -!- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL REGULATORS.

Db	487	ProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu	506
QY	1153	CAAAACAACGGCAGCTCTGATATACAGAAATACTACCGTAACCTCAAAATGTAAACCAAGAT	1212
Db	507	GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp	526
QY	1213	GTAGCGCTGGTGCAGTTCTTGTCCCTGTGCTGTGTCATGTGCTCTCACTACTCTCATCTC	1272
Db	527	ValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu	546
QY	1273	ATATTAGTCTGCTTGGCACTCGGAGAAACAGAAAGAAAAAANAACAGGACCACTATGAC	1332
Db	547	IleLeuValCysAlaThrPHisTrPArgAsnArgLysLysThrGluGlyThrThrAsp	566
QY	1333	TTACCTTACTGGACCGCGGCAAGTTGGTGGAAAGAAATGAAGCAGGTTCCTCTCGCAAAA	1392
Db	567	LeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlaLys	586
QY	1393	GCAGTGGACCATGAGGAACCCCAAGTTTCGTGTAAAGCAGCAGCAAGTTAATCACTGAGT	1452
Db	587	AlaValAspHisGluGluThrProValArgTyrSerSerGluValAsnHisLeuSer	606
QY	1453	CCAAGAGAAGTACCACACAGTCTCGACGCTGACTCTCCAGAGTATGCTCAGGCACCTGGTA	1512
Db	607	ProArgGluValThrThrValLeuGlnAlaAspSerAlaGluThrAlaGlnProLeuVal	636
QY	1513	GGAGGAATTTGGTGACACTTCATCAAGAATCTACCTTTAAACCAGAAAGAAAGAA	1572
Db	627	GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGlyLysGlu	646
QY	1573	GCAGGCTATGCAGACCTAGATCTTTACAACCTCACCGAGGCAAGGTTTATCATGCTCAT	1632
Db	647	AlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyr	666
QY	1633	GCTGAACCACTCCCAATTACGGGGCTCAGTATGCAACCCCAATCATCATGGACATGTCA	1692
Db	667	AlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMetAspMetSer	686
QY	1693	GGCACCCACACAACCTTCAGTTGGTCAGCCCTCCACATCCACTTTCACAGGCTACGGGAAC	1752
Db	687	GlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsn	706
QY	1753	CAACCTCCCCCACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGACAGCTGCTCC	1812
Db	707	GlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSer	726
QY	1813	TCAGCCCGAGCCCGATGATATACCCCGAAGCTGGGAAGCCAGGTCTACCTGCCCCAGAC	1872
Db	727	SerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAsp	746
QY	1873	GAATTGTGTACCAGTGGCCACAGACACACAAGAAGTATCAGAGCAGGAAGGATGGG	1932
Db	747	GluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGly	766
QY	1933	GAATGTGATGTTTTTAAAGAAATCTTT	1959
Db	767	GlucCysAspValPheLysGluIleLeu	775
RESULT 3			
ID	Q912V2	PRELIMINARY;	PRT; 769 AA.
AC	Q912V2;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DR	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Endothelial and smooth muscle cell-derived neuropilin-like protein.		
DE	EC		
GN	ESDN.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=101116;		
RN	[1]		

[illegible]

Db 250 GlnIleSerValValIleSerLysGlyThrProTyrTyrGluSerSerLeuAlaAsnAsn 269
 QY 451 GTCATCTGCTGGGGGACACTTATCTACAGTCTTTTACATTTAAAGACAAGTGGATGT 510
 Db 270 ValThrSerMetValGlyTyrLeuSerThrSerLeuPheThrPheLysThrSerGlyCys 289
 QY 511 TATGAACACTGGGGATGGAGTCTGGTGTATCCGGATCTCTCAATAACAGCATCATCT 570
 Db 290 TyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSer 309
 QY 571 GTCTGGAGTGGACTACACACAGGCAAGCAAGCAAGTGGAAACCCCAAAAAGCCAGG 630
 Db 310 ValLeuGluTrpThrAspHisMetGlyGlnGluAsnSerTrpLysProGluLysAlaArg 329
 QY 631 CTGAAAAAAGCTGACCGCCCTGGGCTGCTTTCCTACTGATGAATACCATGTTTCAA 690
 Db 330 LeuArgLysProGlyProProTrpAlaAlaPheAlaThrAspGluHisGlnTrpLeuGln 349
 QY 691 ATAGATTTGAATAGGAAAGAAATACAGGCATTATACCATCTGGATCCACCATGGTG 750
 Db 350 IleAspLeuAsnLysGluLysIleThrGlyIleValThrThrGlySerThrLeuIle 369
 QY 751 GAGCACAATTACTATGCTGCTGCTACAGAACTCTGACAGTATGATGGCGCAAAATGG 810
 Db 370 GluHisAsnTyrTyrValSerAlaTyrArgValLeuTyrSerAspGlyGlnLysTrp 389
 QY 811 ACTGTGTACAGAGCCTGGTGTGGACAGATAGATATTTCAAGGAACAACAGATAT 870
 Db 390 ThrValTyrArgGluProGlyAlaAlaGlnAspLysIlePheGlnGlyAsnLysAspTyr 409
 QY 871 CACAGGATGTGCGTAATAACTTTTGGCCACCAATTTATGCACCTTTTATTAGAGTGAAT 930
 Db 410 HisLysAspValArgAsnAsnPheLeuProProIleLeuAlaArgPheIleArgValAsn 429
 QY 931 CCTACCCCAATGGCAGCAAAATGCCATGAAATGGAGCTGCTCGGATGTCAGTTTATT 990
 Db 430 ProValGlnTrpGlnGlnLysIleAlaMetLysValGluLeuLeuGlyCysGlnPheThr 449
 QY 991 CCTAAGTCTGCTCCAAAATTACTCAACTCCACTCCCTCCGGAACAGCAATGACCTC 1050
 Db 450 LeuLysGlyArgLeuProLysLeuThrGlnProProProProArgAsnSerAsnLeu 469
 QY 1051 AAAAACAATACAGCCCTCCAAAATAGCCAAAGTCTGCTGCCCAAAATTTACGCAACA 1110
 Db 470 LysAsnThrThrValHisProLysLeu-----GlyArgAlaProLysPheThrGlnAla 487
 QY 1111 CTACAACCTCGCAGTAGCAATGAAATTTCTCCACAGACAGCAACAACATGCCAGTCT 1170
 Db 488 LeuGlnProArgSerArgAsnAspLeuProLeuLeuProAlaGlnThrThrAlaThrPro 507
 QY 1171 GATATCAAAATACTACCTGAACTCAATGTAAACCAAGATGTAGCCTGGCTGCAGTT 1230
 Db 508 AspValLysAsnThrThrValThrProSerValThrLysAspValAlaLeuAlaVal 527
 QY 1231 CTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1290
 Db 528 LeuValProValLeuValMetAlaLeuThrThrLeuIleLeuIleValValCysAlaTrp 547
 QY 1291 CACTGGAGAAACAGAAAGAAAAAAGTGAAGGCACCTATGACTTACCTTACTGGGACGG 1350
 Db 548 HisTrpArgAsnArgLysLysAlaGluGlyThrTyrAspLeuProHisTrpAspArg 567
 QY 1351 GCAGTGTGGTGAAGAGGAATGAGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGAA 1410
 Db 568 AlaGlyTrpTrpLysGlyValLysGlnLeuLeuProAlaLysSerValGluHisGlu 587
 QY 1411 ACCCAGTTCGCTATAGCAGCAGCAAGTAAATCACTGAGTCCCAAGAGAGTCAACACA 1470
 Db 588 ThrProValArgTyrSerAsnSerGluValSerHisLeuSerProArgGluValThrThr 607
 QY 1471 GTGCTGCAAGCTCTGCAAGTATGCTCAGCCACTGGTAGGAGGAATGTTGGTACA 1530
 Db 608 ValLeuGlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThr 627

QY 1531 CTTTCATCAAGATCTACCTTTTAAACCAAGAGAAAGAGCAGGCTATATCAGACCTA 1590
 Db 628 LeuHisGlnArgSerThrPheLysProGluGluGlyLysGluAlaSerTyrAlaAspLeu 647
 QY 1591 GATCCTTACAACCTCACACAGGCGAGGAAGTTTATCATGCTATGCTGAACCACTCCCAATT 1650
 Db 648 AspProTyrAsnAlaProValGlnGluValTyrHisAlaTyrAlaGluProLeuProVal 667
 QY 1651 ACGGGGCTGAGTATGCAACCCCAATCATGTCATGTCAGGACACCCCAACACTTCA 1710
 Db 668 ThrGlyProGluTyrAlaThrProIleValMetAspMetSerGlyHisSerThrAlaSer 687
 QY 1711 GTTGGTCAGCCCTCCACATCCACTTTCAGGCTACGGGGAACCACTCCCTCCACTAGT 1770
 Db 688 ValGlyLeuProSerThrSerThrPheArgThrAlaGlyAsnGlnProProAlaLeuVal 707
 QY 1771 GGAATTAACAATACACTTCTCTCCAGGACTGACAGCTGCTCCTCAGCCAGCCAGTAT 1830
 Db 708 GlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerGlyGlnAlaGlnTyr 727
 QY 1831 GATACCCCGAAGCTGGGAAGCCAGTCTACCTGCCCCAGACGAAATTTGGTGTACCAAGT 1890
 Db 728 AspThrProLysGlyGlyLysProAla---AlaAlaProGluGluLeuValTyrGlnVal 746
 QY 1891 CCACAGACACACAAAGTATCAGGACGAGGAGGATGGGGAATGTGATCTTTTAAA 1950
 Db 747 ProGlnSerThrGlnGluAlaSerGlyAlaGlyArgAspGluLysPheAspAlaPheLys 766
 QY 1951 GAAATCCTT 1959
 Db 767 GluThrLeu 769
 RESULT 4
 Q912V3 PRELIMINARY: PRT; 769 AA.
 ID Q912V3
 AC Q912V3;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
 DE ESDN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR.
 RX PubMed=11447234;
 RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
 RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
 RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
 RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
 RT Up-regulated after Vascular Injury";
 RL J. Biol. Chem. 276:34105-34114(2001).
 CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 DR EMBL; AF387548; AAL30179.1; -;
 DR MGD; MGI:1920629; Esdn.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58_C.
 DR InterPro; IPR004043; LCCL_dom.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF00754; F5_F8_type_C; 1.
 DR PROSITE; PS01180; CUB; 1.
 KW DNA-binding; Transcription regulation.
 SQ SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;
 Alignment Scores: 1.41e-244 Length: 769
 Pred. No.: 2950.50 Matches: 559
 Score:

Db	370	GluHisSerTyrTyrValSerAlaTyrArgValLeuTyrSerAspGlyGlnArgTyr	389
Qy	811	ACTGTGTACAGAGCGCTGTTGGGAGCAAGATAAGATATTTCAAGAGAAACAAGATTAT	870
Db	390	ThrValTyrArgGluProGlyValAspGlnAspLysPheGlnGlyAsnLysAspTyr	409
Qy	871	CACCAGGATGTGGTAACTATTTTCCACCACAAATATTGCACGTTTTATTAGAGTGAAT	930
Db	410	HisLysAspValArgAsnAsnPheLeuProIleAlaArgPheIleArgValAsn	429
Qy	931	CCTACCACAAATGGCAGCAGAAAAATGGCATGAAATYGAGAGCTCTCGATGTCAGTTATT	990
Db	430	ProValGlnTyrGlnGlnLysIleAlaMetLysValGluLeuLeuGlyCysGlnPheThr	449
Qy	991	CCTAAAGSTCGTCTCTCAAAACTTACTCAACCTCCACCTCTCGAAGCAGCAATGACCTC	1050
Db	450	LeuLysGlyArgLeuProLysLeuThr-----ProProProArgAsnGlyAsnAsnLeu	467
Qy	1051	AAAAACACTACAGCCCTCCAAAAATAGCCAAAGTGTGCCCCCAAAATTTACGCAACCA	1110
Db	468	ArgAsnThrThrAlaArgProLysLeuGlyLysGlyArgAlaProLysPheThrGlnVal	487
Qy	1111	CTACAACCTCGCAGTACCAATGAATTTCTCGCAGCAGACAGAAACAACACTGCCAGTCT	1170
Db	488	LeuGlnProArgSerArgAsnGluLeuProValGlnProAlaGluThrThrThrPro	507
Qy	1171	GATATCGAANAATACTACCGTTAACTCCAATGTAAACCAAGATGTAGCGCTGGCTGCATT	1230
Db	508	AspIleLysAsnThrThrValThrProSerValThrLysAspValAlaLeuAlaVal	527
Qy	1231	CTTGTCTCTGCTGGTCACTGGTCTCACTACTCTCATTTCTCATATTAGTGTGTCTGG	1290
Db	528	LeuValProValLeuValMetAlaLeuThrThrLeuIleLeuLeuValCysAlaTyr	547
Qy	1291	CACTGGAGAACAGAAAGAAAAAACAAGCACTATGACTTACTTACTGGAGCCGG	1350
Db	548	HisTyrArgAsnArgLysLysThrGluGlyAlaTyrAspLeuProHisTyrAspArg	567
Qy	1351	GCAGGTTGGTGAAGGAATGAAGCAGTTCTTCTCGAAAGACAGTGGACCATCAGGAA	1410
Db	568	AlaGlyTyrTyrLysGlyMetLysGlnLeuLeuProAlaLysSerValAspHisGluGlu	587
Qy	1411	ACCCCACTTCGCTATACGACGCGAAGTTAATCACTGAGTCCAAAGAGAAGTCAACACA	1470
Db	588	ThrProValArgTyrSerThrSerGluValSerHisLeuSerAlaArgGluValThrThr	607
Qy	1471	GTGCTCGAGGCTGACTCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTGACA	1530
Db	608	ValLeuGlnAlaaspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThr	627
Qy	1531	CTTCATCAAGATCTACCTTTAAACAGAGAGGAAAGAGAGCGCTATGCGACCTTA	1590
Db	628	LeuHisGlnArgSerThrPheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeu	647
Qy	1591	GATCCTTTACAACCTCACAGGGCAGGAAGTTTATCATGCTATGCTGAACCACTCCCAAT	1650
Db	648	AspProTyrAsnSerProMetGlnGluValTyrHisAlaTyrAlaGluProLeuProVal	667
Qy	1651	ACGGGGCTGAGTATGCAACCCCAATCATCATGAGCATGTGAGGCGACCCCAACAATTCA	1710
Db	668	ThrGlyProGluTyrAlaThrProIleValMetAspMetSerGlyHisProThrAlaSer	687
Qy	1711	GTGTGTCAGCCCTCCACATCCACTTTCAAAGGCTACGGGGAACCAACCTCCCCCACTAGT	1770
Db	688	ValGlyLeuProSerThrSerThrPheLysThrAlaGlyThrGlnProHisAlaLeuVal	707
Qy	1771	GGAACTTTACAATACACTTCTCTCCAGCACTGACAGCTGCTCTCAGCCCGCCCGCAGTAT	1830
Db	708	GlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSerGlyGlnAlaGlnTyr	727
Qy	1831	GATACCCCGAAGCTGGGAAGCAGGCTCTACCTGCCCCAGACGAATGGTGTCACAGGTG	1890
Db	728	AspThrProLysGlyLysLys----SerAlaAlaThrProGluGluLeuValTyrGlnVal	746

QY	397	AGTGTGTTAAATTAGTAAAGGTATTCCTCTATTATGAAAGTTCTTTGGCTAACAAAGTCACA	456
Db	166	AsnLeuLeuGlnSerLysGlyIleSerHisGlyGluGlyLeuLeuAlaAsnGlyValLeu	185
QY	457	TCTGTGGTCGGACACTATCTACAGCTCTTTTACATTTTAAGCAACAGTGGATGTTATGCA	516
Db	186	SerArgHisGlySerLeuSerGlnLysArgPheLeuPheThrThr	200
QY	517	ACACTGGGATGGAGTCTGGTGTGATCGCGATCCTCAATAACAGCATCATCTGTGCTG	576
Db	200	-----	200
QY	577	GAGTGGACTGACCACAGGGCAAGAGAACAGTGTGSAACCCAAAAAGCCAGGCTGAAA	636
Db	200	-----	200
QY	637	AAACCTGGACCGCTTGGCTGCTTTTGGCCACTGATGATACCAGTGGTTACAAATAGAT	696
Db	200	-----	200
QY	697	TTGAATAAGGAAAAAGAAAAATAACAGGCATTATAACCACCTGATCCACCATGTGTGGAGCAC	756
Db	200	-----	200
QY	757	AATTACTATGTCTGCCCTACAGAACTCTGTACAGTGAATGGCGAGAATGGACTGTG	816
Db	200	-----	200
QY	817	TACAGAGCGCTGGTGTGGAGCAAGATAAGATAATTTCAGAGAAACAAGATTATCACCAG	876
Db	201	-----ProGlyMet-----	203
QY	877	GATGTGCGTAATAACTTTTTTGGCCACCAATTATTGCAGCTTTATTAGAGTGAATCCTACC	936
Db	203	-----	203
QY	937	CAATGGCAGCAGAAAATTGCCATGAAATGGAGCTGCTCGGATGTCAGTTTATTCCTAAA	996
Db	203	-----	203
QY	997	GGTGTCTCCAAAACCTTACTCAACCTCCACCTCCTCGGAACGAATGACCTCAAAAAC	1056
Db	203	-----	203
QY	1057	ACTAGACCCCTCCAAAAATAGCCAAAGCTGTCGCCCAAAATTACGCAACCACTACAA	1116
Db	203	-----	203
QY	1117	CCTCGCAGTAGCAATGAATTTCTCGACAGACAGAACAACTGCCAGTCCCTGATATC	1176
Db	203	-----	203
QY	1177	AGAAATACTACCGTAACCTCCAAATGTAAACAAAGATGTAGCGCTGGCTGCAGTTCCTTGC	1236
Db	204	-----AsnIleThr--ThrValAlaIleProSerValIle--	214
QY	1237	CCTGTGCTGTCATGGTCTCTACTACTCTCATTCATATTAGTGTGTGCTGGCACTGG	1296
Db	215	---PheIleAlaLeuLeuThrGlyMetGlyIlePheAlaIleCys---	229
QY	1297	AGAAACAGAAAGAAAAAACTCAAGCGACCTATGACTTACCTTAC	1344
Db	230	-----ArgLysArgLysLysGlyAsn-----ProTyrValSerAlaAspAla	244
QY	1345	GACCGCGCAGGTTGGTGGAAAGAAATGAAGCAGTTTCTTCCTGCAAAAGCAGTGGACCAT	1404
Db	245	GlnLysThrGlyCysTrpLysGlnIleLysTyr-----	255
QY	1405	GAGAAACCCCA--GTTTGGCTATAGCAGCACCGAA-----GTTAATCACTGAGTCCA	1455
Db	256	-----ProPheAlaArgHisGlnSerThrGluPheThrIleSerTyrAspAsnGlu	272

Qy	1456	AGAGAACTCAC	-----ACAGTGTGTCAGGCTGACTCTGCAGACGATGATGCTACGCCA	1506
Db	273	LysGluMetThrGlnLysLeuAspLeuIleThrSerAspMetAlaAspTyrGlnGlnPro	292	
Qy	1507	CTGTGTAGGAGGATTTGTTGGTACACTTCATCAAGACTCTACCTCTTAAACCA	1557	
Db	293	LeuMetIleGlyThrGlyThrValAlaAArgLysGlySerThrPheArgProMetAspThr	312	
Qy	1558	-----GAAGAA-----GGAAAGAGGAGCGGTATGCAGACCTAGATCCTTAC	1599	
Db	313	AspThrGluGluValArgValAsnThrGluAlaSerGlyHis	327	
Qy	1600	AACCTCACCAGGGCAGGAAGTTTATCATGCTATGCTGACCACTCCCAATTACGGGGCT	1659	
Db	328	AspCysProHisArgProGlyArgHisGluTyrAlaLeuProLeuThrHisSerGluPro	347	
Qy	1660	GAGTATCAACCCCAATCATC-----ATGGACATG	1699	
Db	348	GluTyrAlaThrProIleValGluAArgHisLeuLeuArgAlaHisThrPheSerThrGln	367	
Qy	1690	TCAGGGCAGCCCACTTCAGTTGTGTGACGCTCCACATCCACTTCAAGGCTACGGGG	1749	
Db	368	SerGlyTyrArgValProGly---ProArgProThrHisLysHisSerHisSerSerGly	386	
Qy	1750	AACCAACCTCCCCCACTAGTGGGAACCTTACAATACACTTCTCCAGGACTGCACAGTGC	1809	
Db	387	GlyPheProProAlaThrGlyAlaThrGlnValGluSerTyrGlnArgProAlaSerPro	406	
Qy	1810	TCCTCAGCCAGGCCCCAGTATGATACCCGGAAGCTGGGAAGCAGGTCTACCTCCGCCA	1859	
Db	407	LysProValGlyGlyTyrAspLysPro---AlaAlaSerSerPheLeuAspSerArg	425	
Qy	1870	GACCAATTTGTGTACCAAGTGCACAGACACACAAGAAGTATCAGGACGAGGAAGGAT	1929	
Db	426	Asp-----ProAlaSerGlnSerGlnMetThrSerGlyGlyAspAsp	439	
Qy	1930	GGG 1932		
Db	440	Gly 440		
RESULT 8				
ID	Q8R327	PRELIMINARY;	PRT;	432 AA.
AC	Q8R327;			
DT	01-JUN-2002 (TEMBLrel. 21, Created)			
DT	01-JUN-2002 (TEMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TEMBLrel. 21, Last annotation update)			
DE	Similar to RIKEN cDNA 463143K11 gene.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC026771; AAH26771.1; -			
SQ	SEQUENCE 432 AA; 46798 MW; AD3FB42B24CD4917 CRC64;			
Alignment Scores:				
Pred. No.:	7,72e-27	Length:	432	
Score:	414.00	Matches:	149	
Percent Similarity:	32.89%	Conservative:	72	
Best Local Similarity:	22.17%	Mismatches:	155	
Query Match:	11.63%	Indels:	296	
DB:	11	Gaps:	19	

QY 1558 -----GAGAA-----GGAAGACAGCGCTATGACAGCTTATGATCTTAC 1599
 |||||
 Db 356 AspThrGluGluValAlaArgValAsnThrGluAlaSerGlyHis-----Tyr 370
 |||||
 QY 1600 AACCCACAGGGGCGAGAGATTATCATGCCATGCTGACACATCCCATTTACGGGGCT 1659
 |||||
 Db 371 AspGlySerProHisArgProGlyArgGlyGluValAlaLeuProLeuThrHisSerGluPro 390
 |||||
 QY 1660 GAGTATGCAACCCCAATC-----ATGACATG 1689
 |||||
 Db 391 GluThrAlaThrProIleValGluArgHisIleuLeuArgAlaHisThrPheSerThrGln 410
 |||||
 QY 1690 TCAGGACGACCCGACACTTCAGTGGTCCAGCCCTCCACATCCACTTTCAGGCTACGGGG 1749
 |||||
 Db 411 SerGlyThrArgValProGly---ProArgProThrHisLysHisSerHisSerSerGly 429
 |||||
 QY 1750 AACCACTCCGACATGAGGAACTTACATACACTTCTCTCCAGGCTGACAGCTGC 1809
 |||||
 Db 430 GlyPheProProAlaThrGlyAlaThrGlnValGluSerThrGlnArgProAlaSerPro 449
 |||||
 QY 1810 TCCACGACCCGACGCGATGATACCCCGAAGCTGGAAGCCAGGTCTACGCGCCCA 1869
 |||||
 Db 450 LysProValGlyGlyGlyTyrAspLysPro---AlaAlaSerSerPheLeuAspSerArg 468
 |||||
 QY 1870 GACCAATTGGTGTACAGGTGCGACAGACACACAGAGATATCAGGACAGAGGAT 1929
 |||||
 Db 469 Asp-----ProAlaSerGlnSerGlnMetThrSerGlyLysAsp 482
 |||||
 QY 1930 GGG 1932
 |||||
 Db 483 G1Y 483

RESULT 10

Q90X47 PRELIMINARY: PRT: 2119 AA.

AC Q90X47: 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SC:b22015.3 (novel protein similar to vertebrate coagulation factor V and VIII).
 GN SC:B22015.3.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 CC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lloyd D.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL590146; CAC94896.1; -;
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR000421; FA58_C.
 DR Pfam: PF00394; Cu-oxidase; 2.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR PROSITE: PS01285; FA58C_1; UNKNOWN_2.
 DR PROSITE: PS01286; FA58C_2; UNKNOWN_1.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_3.
 SQ SEQUENCE 2119 AA; 240643 MW; DCOE806FFA8761E6 CRC64;

Alignment Scores:

Pred. No.: 6,31e-22 Length: 2119
 Score: 359.00 Matches: 72
 Percent Similarity: 62.20% Conservative: 30
 Best Local Similarity: 43.90% Mismatches: 54
 Query Match: 10.09% Indels: 8
 Gaps: 4

US-10-060-830-1113 (1-1962) x Q90X47 (1-2119)

QY 505 GGATGTTATGCAACACTGGGATGGAGTCTGCTGATCGCGGATCCTCAATAACAGCA 564

Db 1660 GlyCysSerValProLeuGlyMetGlyIleuLysGlyLeuIleAspSerLysThrAla 1979
 |||||
 QY 565 TCATCTGTGCTG---GAGTGGACTGACCAACACAGGCGCAAGAGAACGTTGGAAACCAAA 621
 |||||
 Db 1980 SerSerValAlaSerAsnTrp-----TyrSerGlyGln-----TrpHisProTrp 1994
 |||||
 QY 622 AAAGCAGGCTGAAACAACTGACCG-----CCTGGGCTCTTTGCCACTGATGAA 675
 |||||
 Db 1995 TyrAlaArgLeuAsnLysGlnGlyThrAlaAsnAlaTrpGlnAlaLysAsnAspIle 2014
 |||||
 QY 676 TACCGTGTTCACAAATTCATTCATTAAGCAAAAGAAATATACGGCTATATACCACT 735
 |||||
 Db 2015 GlnProTrpIleGlnValGluIleuLysGluValLysLysThrGlyIleValThrGln 2034
 |||||
 QY 736 GGATGACACAGTGGGACGACCAATACATATGCTGCTGCACAGAAATGCGTACAGTGT 795
 |||||
 Db 2035 GlyAlaLysSerMetGlyAsnGluMetPheValArgSerThrIleGluIleuLysSerGlu 2054
 |||||
 QY 796 GATGGGCAAAATGACATGCTGTACAGAGAGCCGTGGTGGAGCAAGATATATTTCAA 855
 |||||
 Db 2055 AspGlyArgArgTrpMetLysThrThrAspAspPtyrGluGlnLysLeuPheGln 2074
 |||||
 QY 856 GGAACAAAGATTATCCACAGAGATGCGTAAATCTTTCACAGCAATATTCACAGT 915
 |||||
 Db 2075 GlyAsnThrAspAsnAsnGlyGlnIleLysAsnThrIleTyrProProIlePheSerArg 2094
 |||||
 QY 916 TTTATATAGCAATCCATCCATGCGACAGACAGAAATGCGCAAAATGCGATGCTGC 975
 |||||
 Db 2095 PheIleArgGlyIleLeuProLysGlnTrpGlnLysSerValThrMetArgIleIleuLeu 2114
 |||||
 QY 976 GGATGCTCAGTTT 987
 |||||
 Db 2115 GlyCysAspPhe 2118

RESULT 11

O88783 PRELIMINARY: PRT: 2183 AA.

AC O88783: 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Murine coagulation factor V.
 GN F5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yang T.L., Cai J., RehmTulla A., Yang A., Mousalli M., Kaufman R.J.,
 Ginsburg D.;
 RL "The structure and function of murine factor V and its inactivation by
 protein C.";
 RL Blood 91:4593-4599(1998).
 CC -1- SIMILARITY: CONTAINS 2 F5/F8 TYPE C DOMAINS.
 DR EMBL: U52925; AAC99553.1; -;
 DR HSP: P12259; ICDT.
 DR MSD: MGI:88382; F5.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR000421; FA58_C.
 DR Pfam: PF00394; Cu-oxidase; 3.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR SMART: SM00231; FA58C_2.
 DR PROSITE: PS01285; FA58C_1; 2.
 DR PROSITE: PS01286; FA58C_2; 2.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 2.
 SQ SEQUENCE 2183 AA; 247228 MW; BFO8AA723F60317 CRC64;

Alignment Scores:

Pred. No.: 1.27e-21 Length: 2183
 Score: 355.50 Matches: 79


```

Db 365 AsnaAlatrrhrserglhHisAsnaSpolnsrIntrpreuGlInValAspleuVal 384
QY 706 GAAAGAAATTAACAGCATTTATTAACAGTCAGTCCAGCATGTCAGACAAATTAATAT 765
Db 385 ProThrlysalnrlhrglyllelthrglnlValAlaLysAsrPheglYhIsValGlnPhe 404
QY 766 GTGTCTGCCATACAGAAATCCGTACAGTGAATGGGAGAAATGAGTGTGTACAGAG 825
Db 405 ValGlySerlyrlylsleuAlaTySerAsnaSpolnysrlIntrprthValTyGlnAsp 424
QY 826 CTTGGTGTGGAGCAAGATTAATTCAGAGAAACAAGATTTATCACCAGAGATGGCT 885
Db 425 GlulysGlnarGlysaSprlySvalPheGlnGlyAsnPhesAsnaSprThnIsArgLys 444
QY 886 AATTAATTTTGGCCACCAATTAATTCAGCTTTATTAAGTGAATGATCCATACCAATGAG 945
Db 445 AsnValIleAsrProrProIleTyAlaArgHisIleArgIleAsrProTprSerTrpYr 464
QY 946 CAGAAATTCCTCATGAATATGAGCTCTCGGATGT 981
Db 465 GlyArgIleThrLeuArgSerGluLeuGlyCys 476

```

RESULT 13

035474 PRELIMINARY; PRT; 480 AA.

```

AC 035474, 035475;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Integrin-binding protein DEL1 precursor.
GN EDL3 OR DEL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid-10090;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RC TISSUE-EMBRYO;
RX MEDLINE-98083109; PubMed-9420328;
RA Hidai C., Zupancic T., Penta K., Mikhail A., Kawana M.,
RA Quertermous E.E., Aoka Y., Snodgrass M., Matsui Y., Platika D.,
RA Auerbach R., Hogan B.L.M., Snodgrass R., Quertermous T.;
RT *Cloning and characterization of developmental endothelial locus-1: an
RT embryonic endothelial cell protein that binds the alphabeta3 integrin
RT receptor.
RT Genes Dev. 12:21-33(1998).
RL -1- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
RL INTERACTION WITH THE ALPHA-V BETA3 INTEGRIN RECEPTOR. INHIBITS
RL FORMATION OF VASCULAR-LIKE STRUCTURES. MAY BE INVOLVED IN
RL REGULATION OF VASCULAR MORPHOGENESIS OR REMODELING IN EMBRYONIC
RL DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS; LONG ISOFORM (SHOWN HERE) AND
CC SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ANGIOBLASTS AND EARLY ENDOTHELIAL
CC CELLS. BY EMBRYONIC DAY 13.5, ALSO EXPRESSED IN A RESTRICTED GROUP
CC OF NON-ENDOTHELIAL CELLS INCLUDING CHONDROCYTES AND RETINAL
CC NEURONS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE EMBRYO FROM DAY 7. AFTER DAY
CC 15.5, EXPRESSION DECREASES AND DISAPPEARS COMPLETELY BY THE TIME
CC OF BIRTH.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC EMBL: AF031524; AAB86585.1; -
CC EMBL: AF031525; AAB86586.1; -
CC HSSP: P00750; 1TPG.
CC MGD: MGI:1329025; Edl13.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000742; EGF-2.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR000421; FA58_C.

```

```

DR Pfam; PF00008; EGF_3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_1like; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_3.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW EGF-like domain; Alternative splicing; Signal; Developmental protein;
KW Cell adhesion; Repeat; Vascularization.
FT SIGNAL 1
FT CHAIN 17 480
FT DOMAIN 26 59
FT DOMAIN 78 116
FT DOMAIN 123 154
FT DOMAIN 161 311
FT DOMAIN 322 473
FT SITE 96 98
FT DISULFID 26 37
FT DISULFID 31 48
FT DISULFID 50 59
FT DISULFID 78 89
FT DISULFID 83 105
FT DISULFID 107 116
FT DISULFID 158 314
FT DISULFID 301 305
FT DISULFID 319 476
FT VARSPPLIC 218 221
FT VARSPPLIC 222 480
SQ SEQUENCE 480 AA; 53740 MW; 4CD91EFE926171AD CRC64;

```

Alignment Scores:

```

Pred. No.: 1.68e-21 Length: 480
Score: 352.00 Matches: 75
Percent Similarity: 59.30% Conservative: 27
Best Local Similarity: 43.60% Mismatches: 66
Query Match: 9.89% Indels: 4
DB: Gaps: 2

```

US-10-060-830-1113 (1-1962) x 035474 (1-480)

```

QY 472 TTATCTACAACTCTTTTATTAAGCAAGATGATGTTATGAGACCTGGGATGAG 531
Db 307 LeuArgMetGluLeuLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMetLys 326
QY 532 TCTGTGTGATCGCGGATCTCAATTAACAGCATCTGTCTGGAGTGGACTGACAC 591
Db 327 SerGlyHisIleGlnAspTyGlnIleThrAlaSerSerValIlePheArgThrLeuAsnMet 346
QY 592 ACAGGCAAGAGACAGTGGGAAACCCAAACCAAGGCTGAAACCACTGACCG-- 648
Db 347 -----AspMetPheThrTrpGluProArgGlyAlaArgLeuAspLysGlnGlyVal 364
QY 649 ---CCTGGCGCTCTTTGCCATGATGAATCAAGTGGTATCAATAGATTGAATAG 705
Db 365 AsnaAlatrrhrserglhHisAsnaSpolnsrIntrpreuGlInValAspleuVal 384
QY 706 GAAAGAAATTAACAGCATTTATTAACAGTCAGTCCAGCATGTCAGACAAATTAATAT 765
Db 385 ProThrlysalnrlhrglyllelthrglnlValAlaLysAsrPheglYhIsValGlnPhe 404
QY 766 GTGTCTGCCATACAGAAATCCGTACAGTGAATGGGAGAAATGAGTGTGTACAGAG 825
Db 405 ValGlySerlyrlylsleuAlaTySerAsnaSpolnysrlIntrprthValTyGlnAsp 424
QY 826 CTTGGTGTGGAGCAAGATTAATTCAGAGAAACAAGATTTATCACCAGAGATGGCT 885
Db 425 GlulysGlnarGlysaSprlySvalPheGlnGlyAsnPhesAsnaSprThnIsArgLys 444
QY 886 AATTAATTTTGGCCACCAATTAATTCAGCTTTATTAAGTGAATGATCCATACCAATGAG 945

```

Qy	904	ATTTATTGACGGTTTTTATTAGATGAATCTACCCAATGCCACACAAAATTGCCATGAA	963
Dd	2196	IleIleSerApgheIleaRgValIIeProLysThrrTPAsnGInSerIleAlaLeuArg	2215
Qy	964	ATGAGCGTCGTGCGATGT 981	
Dd	2216	LeuGlutLeuPheGlyCys 2221	
RESULT 15			
O18806		PRELIMINARY; PRT; 2343 AA.	
ID	O18806		
AD	O18806;		
DT	01-JAN-1998 (TREMBLrel_05, Created)		
DT	01-JAN-1998 (TREMBLrel_05, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel_15, Last annotation update)		
DE	Factor VIII.		
DN	F8.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
FX	NCBL_TaxID=9615;		
RN	[1]		
RS	SEQUENCE FROM N.A.		
RP	TSUBE=LIVER;		
RC	Tissue=Liver;		
RA	Cameron C., Nottley C., Hoyle S., McGlynn L., Hough C., Kamisie S.,		
RA	Giles A., Lilliecrap D.;		
RT	"The canine factor VIII cDNA and 5' flanking sequence.";		
RL	Submitted (NOV-1997) to the EMBL/GenBank/DDBB databases.		
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.		
DR	EMBL; AF016234; AAB87412.1; ..		
DR	HSP; P00451; ICFG		
DR	InterPro; IPR001117; Cu-oxidase.		
DR	InterPro; IPR000421; FA58.C.		
DR	Pfam; PF00394; Cu-oxidase; 3.		
DR	Pfam; PF00754; F5_F8_type_C; 2.		
DR	SMART; SM00231; FA58C; 2.		
DR	PROSITE; PS01285; FA58C.1; 2.		
DR	PROSITE; PS01286; FA58C.2; 2.		
DR	PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.		
SQ	SEQUENCE 2343 AA; 265829 MW; A854FAE571C3B399 CRC64;		
Alignment Scores:			
Pred. No.:	5.39e-19	Length: 2343	
Score:	325.00	Matches: 99	
Percent Similarity:	43.60%	Conservative: 51	
Best Local Similarity:	28.78%	Mismatches: 121	
Query Match:	9.13%	Indels: 73	
DB:	6	Gaps: 11	
US-10-060-830-1113 (1-1962) x O18806 (1-2343)			
Qy	48	GGAGCTGGAGCCCA---GCAGGCAANTACTGTGCTGGGTTCGAAT-----	95
Dd	2055	GlyGlnTrpAlaProLysLeuAlaArgLeuHisrYrSerGlySerIleAsnAlaTrpSr	2074
Qy	96	-----GAACATTCATTAATCAAAGGCAATGAATCACATTCGTGTTTCATGAGTG 149	
Dd	2075	ThrLysaspPropheSerTrpIle-LysValasLeuAlaAlaPrometlleleHsl 2094	
Qy	150	AATCATGTTTCTGGAGCGCGGATTTTTTGGCTCATACTCTGTTATAGATAAACAGATCT	2099
Dd	2094	yIleMetThrGlnGly	
Qy	210	AATFACCTGTTTGGACACATGTCATCCAAATTTTGGAACTGAGTTCAGTAAGTACTGCC 269	
Dd	2100	-----AlaArgGlnLysPheSerSerLeuTyVa 2105	
Qy	270	AGCTGGTGTCTGCTTCTCTTGTCTGAGATATCTGGAACAATTCCTCATGGATATAGGA 329	
Dd	2109	IserGlnPheIlelleMettyrSer----LeuasGlyAsnLysTrphIsertyrargl 2124	
Qy	330	TTCTCGCCATTGTGCATGGCTGTGTGCATGCGAGGAGTAGTGTCAAACACGTGGGGCG 389	

```
Db 2128 yasnSer-----ThrglyThrLeuMetVa 2136
OY 390 CCAATTCAGTGTGTATTAATAGTAAGTATP-----CCCTATTA 428
Db 2136 lPhePheGlyValnValAspSerSerGlyIleLeuHisAsnIlePheAsnProProIleI 2156
OY 429 TGAAGTCTTTGGCTTAACAAGTCA-----TCTGTGGTGGACACTTATCTACAAAG 482
Db 2156 eAlaGlnTyrIleArgLeuHisProThrHisTyrSerIleArgSerThrLeuArgMetG 2176
OY 483 TCTTTTACATTTAAGACAAGTGGATGTATGGAACACTGGGATGGAGTGGTGGTGGAT 542
Db 2176 uLeuLeuGlyCysAspPheAsnSerCysSerMetProLeuGlyMetGlnSerIysAlaI 2196
OY 543 CGGGATCTCTCAATTAACAGCATCATCTGTGTGAGTGAAGTGAACACACAGGGGCAAGA 602
Db 2196 eSerAspAlaGlnIleThrAlaSerSerTyrLeu-----SerSerMetLe 2211
OY 603 GAACAGTTGGAAAAACCCAAAAAGCCAGGCTGAAAAAAGCTGA-----CCGCTTGGGC 656
Db 2211 uAlaThrTyrSerProSerGlnAlaArgLeuHisLeuGlnGlyArgThrAsnAlaIleTrp 2231
OY 657 TGCCTTGCACATGATGAATACAGTGGTGTCAATAGATTGGAATAAGAAAAAAGAAAT 716
Db 2231 gProGlnAlaAsnAsnProLysGlnTyrLeuGlnValAspPheArgLysThrMetLysVa 2251
OY 717 AACAGCATTAATTAACAGTGGATCCACCATGGTGGAGACAAATTAATGATGTGTGCTGA 776
Db 2251 lThrGlyIleThrThrGlnGlyValLysSerLeuLeuIleSerMetTyrValLysGlnPh 2271
OY 777 CAGAAATCCTGTACAGTGAATGATGGGCAAAAAAGTGGATGTGTACAGAGAGCCTGTGGA 836
Db 2271 eLeuIleSerSerSerGlnAspGlnHisAsnTyrThrLeuPheLeuGlnAsnGly----- 2289
OY 837 GCAAGATTAAGATATTTCAGAGAAACAAAGATTATCACCAGATGTGGCTAATTAATT 896
Db 2290 -LysValLysValPheGlnGlnLysAsnArgAspSerSerThrProValArgAsnArgLeuG 2309
OY 897 GCCACCAATTATTCACAGTATTATTAAGTGAATCTTACCCCAATGGCAGCAAAAAATTGC 956
Db 2309 uProProLeuValAlaArgTyrValArgLeuHisProGlnSerThrAlaHisHisIleAl 2329
OY 957 CATGAAAAATGAGCTGCTGGATGTCAAGTTATTCTTAAAGGTGCTCCTCCAAAACTTAC 1016
Db 2329 aLeuArgLeuGlnValLeuGlnCysAsp-----Th 2339
OY 1017 TCAACCTCA 1026
Db 2339 rGlnGlnPro 2342
```

Search completed: January 21, 2003, 09:49:43
Job time : 88.9384 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: January 21, 2003, 09:31:00 ; Search time 13.1166 Seconds
(without alignments)
8802.243 Million cell updates/sec

Title: US-10-060-830-1113
Perfect score: 3559
Sequence: 1 atgcctctgtctctgct.....tttttaagaacacctttga 1962

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US10060830/runat_16012003_091240_26877/app_query.fasta_1.6933
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830 -CGN_1_1_50_erunat_16012003_091240_26877 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_MMJAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	355.5	10.0	2183	3	US-08-746-111-5
2	352	9.9	480	2	US-08-480-229C-10
3	352	9.9	480	2	US-08-659-235C-10
4	352	9.9	513	2	US-08-480-229C-14
5	352	9.9	513	2	US-08-659-235C-14
6	351	9.9	218	1	US-07-607-538C-4
7	351	9.9	218	2	US-08-162-402B-4
8	345	9.7	160	2	US-08-162-402B-14
9	343	9.6	217	1	US-07-607-538C-3
10	343	9.6	217	2	US-08-162-402B-3
11	343	9.6	218	1	US-07-607-538C-2
12	343	9.6	218	2	US-08-162-402B-2

13	343	9.6	387	2	US-08-162-402B-6	Sequence 6, Appli
14	343	9.6	465	2	US-08-162-402B-8	Sequence 8, Appli
15	341	9.6	159	2	US-08-162-402B-12	Sequence 12, Appli
16	330	9.3	321	2	US-08-480-229C-21	Sequence 21, Appli
17	330	9.3	321	2	US-08-659-235C-21	Sequence 21, Appli
18	325	9.1	2343	4	US-09-324-867-2	Sequence 2, Appli
19	323.5	9.1	2304	4	US-09-324-867-4	Sequence 4, Appli
20	322	9.0	2319	1	US-08-212-133A-8	Sequence 8, Appli
21	322	9.0	2319	1	US-08-474-503-6	Sequence 6, Appli
22	322	9.0	2319	2	US-08-670-707A-6	Sequence 6, Appli
23	322	9.0	2319	4	US-09-037-601-6	Sequence 6, Appli
24	322	9.0	2319	4	US-09-315-179-6	Sequence 6, Appli
25	322	9.0	2319	4	US-09-523-656-28	Sequence 28, Appli
26	322	9.0	2319	5	PCT-US94-13200-6	Sequence 6, Appli
27	317	8.9	2351	6	542260-1	Patent No. 542260
28	316	8.9	1438	4	US-09-209-916-1	Sequence 1, Appli
29	316	8.9	1471	1	US-08-683-839B-3	Sequence 3, Appli
30	316	8.9	1661	2	US-08-882-083-2	Sequence 2, Appli
31	316	8.9	1661	2	US-08-558-107-2	Sequence 2, Appli
32	316	8.9	1661	4	US-09-243-539-2	Sequence 2, Appli
33	316	8.9	2332	1	US-07-864-004B-4	Sequence 4, Appli
34	316	8.9	2332	1	US-08-251-937A-4	Sequence 4, Appli
35	316	8.9	2332	1	US-08-212-133A-2	Sequence 2, Appli
36	316	8.9	2332	1	US-08-276-594A-2	Sequence 2, Appli
37	316	8.9	2332	1	US-08-474-503-2	Sequence 2, Appli
38	316	8.9	2332	2	US-08-670-707A-2	Sequence 2, Appli
39	316	8.9	2332	4	US-09-037-601-2	Sequence 2, Appli
40	316	8.9	2332	4	US-09-324-867-3	Sequence 3, Appli
41	316	8.9	2332	4	US-09-315-179-2	Sequence 2, Appli
42	316	8.9	2332	4	US-09-523-656-2	Sequence 2, Appli
43	316	8.9	2332	5	PCT-US93-03275-4	Sequence 4, Appli
44	316	8.9	2332	5	PCT-US94-13200-2	Sequence 2, Appli
45	316	8.9	2351	1	US-08-121-202-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-746-111-5
; Sequence 5, Application US/08746111
; Patent No. 6066778
; GENERAL INFORMATION:
; APPLICANT: Ginsburg, David
; TITLE OF INVENTION: Compositions And Methods For Screening
; TITLE OF INVENTION: Compounds For Anticoagulant Activity
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,111
; FILING DATE: 06-NOV-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2183 amino acids


```
US-08-659-235C-10
; Sequence 10, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,235C
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Polissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0034-999
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 10:
; LENGTH: 480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-659-235C-10
Alignment Scores:
Pred. No.: 8,15e-26 Length: 480
Score: 352.00 Matches: 75
Percent Similarity: 59.30% Conservative: 27
Best Local Similarity: 43.60% Mismatches: 66
Query Match: 9.89% Indels: 4
DB: 2 Gaps: 2
US-10-060-830-1113 (1-1962) x US-08-659-235C-10 (1-480)
QY 472 TTATCTACAGTCTTTTACATTTAACAAGTGGATGTTATGGAACACTGGGATGAG 531
Db 307 LeuArgMetGluLeuLeuGlyCysGluLeuSerGlyCysSerGluProLeuGlyMetLys 326
QY 532 TCTGGTGTGTCGGGATCCTCAATAACAGCATCTCTGTGTGGAGTGGACTGACCAC 591
Db 327 SerGlyHisIleGlnAspTyrGlnIleThrAlaSerSerValPheArgThrLeuAsnMet 346
QY 592 ACAGGGCAAGACACATTGGAACCCCAAAAGCCAGCTGAAACACCTGGACCG--- 648
Db 347 -----AspMetPheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysVal 364
QY 649 ---CCTTGGGCTGCTTTGGCCACTGATGATACAGTGGTTACAAATAGATTTCAATAAG 705
Db 365 AsnAlaTrpThrSerGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuLeuVal 384
QY 706 GAAAGAAAATAACAGGCATTATTAACCACTGGATCCACCATGTTGGAGCACAAATTACTAT 765
Db 385 ProThrLysValThrGlyIleIleThrGlnGlyAlaLysAspPheGlyHisValGlnPhe 404
QY 766 GTGTGCTGTACAGAACTCTGTACAGTGTATGATGCGCAGAAATGGACTGTGTACAGAG 825
Db 405 ValGlySerTyrLysLeuAlaTyrSerAsnAspGlyGluHisTrpMetValHisGlnAsp 424
QY 826 CTTGGTGTGGAGCAAGATAAGATATTTCAGAGAAACAAGAAATATATACCAGGATGTCGT 885
Db 425 GluLysGlnArgLysAspLysValPheGlnGlyAsnPheAspAsnAspThrHisArgLys 444
QY 886 ATAACATTTTGGCCACCAATATTGACAGTGTATTTATAGAGTGAATCTACCAATGGCAG 945
Db 445 AsnValIleAspProIleTyrAlaArgPheIleArgIleLeuProTrpSerTrpTyr 464
QY 946 CAGAAAATGCCATGAATGGAGCTGCTCGGATGT 981
Db 465 GlyArgIleThrLeuArgSerGluLeuLeuGlyCys 476
RESULT 4
US-08-480-229C-14
; Sequence 14, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Polissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0026-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-229C-14
Alignment Scores:
Pred. No.: 8,42e-26 Length: 513
Score: 352.00 Matches: 75
Percent Similarity: 58.72% Conservative: 26
Best Local Similarity: 43.60% Mismatches: 67
Query Match: 9.89% Indels: 4
DB: 2 Gaps: 2
US-10-060-830-1113 (1-1962) x US-08-480-229C-14 (1-513)
QY 472 TTATCTACAGTCTTTTACATTTAACAAGTGGATGTTATGGAACACTGGGATGAG 531
Db 472 TTATCTACAGTCTTTTACATTTAACAAGTGGATGTTATGGAACACTGGGATGAG 531
```

```

; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-659-235C-14

Alignment Scores:
Pred. No.: 8,42e-26 Length: 513
Score: 352.00 Matches: 75
Percent Similarity: 58.72% Conservatives: 26
Best Local Similarity: 43.60% Mismatches: 67
Query Match: 9.89% Indels: 4
DB: 2 Gaps: 2

US-10-060-830-1113 (1-1962) x US-08-659-235C-14 (1-513)
QY 472 TTATCTACAAAGTCTTTTACATTAAGACAAGTGGATGTTATGAACACACTGGGATGGAG 531
Db 340 LeuArgMetGluLeuLeuGlyCysSerGluProLeuGlyMetLys 359
QY 532 TCTGGTGTGATCGCGGATCCTCAAAATAACAGCATCATCTGTCTGTGGAGTGGAGTGGAC 591
Db 360 SerGlyHisIleGlnAspGlyGlnIleThrAlaSerSerIlePheArgThrLeuAsnMet 379
QY 592 ACAGGCAAGAGAACAGTGTGAACCCAAAGCCAGCGCTGAAAAAACCCTGGACGC--- 648
Db 380 -----AspMetPheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysVal 397
QY 649 ---CCTGGTGGCTCTTTCCCACTGATGATACCACTGCTTACAAATACATGATGATGAT 705
Db 398 AsnAlaTrpThrSerGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuLeuVal 417
QY 706 GAAAGCAAAATACAGCATTAATACCACTGGATCCACCATGCTGGAGCACAAATTAAT 765
Db 418 ProThrLysValThrGlyIleIleThrGlnGlyAlaLysAspPheGlyHisValGlnPhe 437
QY 766 GTGTCTGCTACAGATCCTGTACAGTGTGATGGCAGAAATGGAGTGTGTACAGAGAG 825
Db 438 ValGlySerTyrlLysLeuAlaTySerAsnAspGlyGluHisTrpThrValTyGlnAsp 457
QY 826 CTTGGTGTGGAGCAAGATATTTCAAGGAAACAAAGATATATCACCAGGATGTCGCT 885
Db 458 GluLysGlnArgLysAspLysValPheGlnGlyAlaAsnAspThrHisArgLys 477
QY 886 AATAACTTTTGGCCACCAATTAATGACAGTGTGATGAGTGAATCCTACCCCAATGCG 945
Db 478 AsnValIleAspProIleTyAlaArgHisIleArgIleLeuProTrpSerTrpTyr 497
QY 946 CAGAAATTCATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 981
Db 498 GlyArgIleThrLeuAlaSerGluLeuLeuGlyCys 509

```

```

RESULT 5
US-08-659-235C-14
; Sequence 14, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quertemous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,235C
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0034-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-659-235C-14

Alignment Scores:
Pred. No.: 8,42e-26 Length: 513
Score: 352.00 Matches: 75
Percent Similarity: 58.72% Conservatives: 26
Best Local Similarity: 43.60% Mismatches: 67
Query Match: 9.89% Indels: 4
DB: 2 Gaps: 2

US-10-060-830-1113 (1-1962) x US-08-659-235C-14 (1-513)
QY 472 TTATCTACAAAGTCTTTTACATTAAGACAAGTGGATGTTATGAACACACTGGGATGGAG 531
Db 340 LeuArgMetGluLeuLeuGlyCysSerGluProLeuGlyMetLys 359
QY 532 TCTGGTGTGATCGCGGATCCTCAAAATAACAGCATCATCTGTCTGTGGAGTGGAGTGGAC 591
Db 360 SerGlyHisIleGlnAspGlyGlnIleThrAlaSerSerIlePheArgThrLeuAsnMet 379
QY 592 ACAGGCAAGAGAACAGTGTGAACCCAAAGCCAGCGCTGAAAAAACCCTGGACGC--- 648
Db 380 -----AspMetPheThrTrpGluProArgLysAlaArgLeuAspLysGlnGlyLysVal 397
QY 649 ---CCTGGTGGCTCTTTCCCACTGATGATACCACTGCTTACAAATACATGATGATGAT 705
Db 398 AsnAlaTrpThrSerGlyHisAsnAspGlnSerGlnTrpLeuGlnValAspLeuLeuVal 417
QY 706 GAAAGCAAAATACAGCATTAATACCACTGGATCCACCATGCTGGAGCACAAATTAAT 765
Db 418 ProThrLysValThrGlyIleIleThrGlnGlyAlaLysAspPheGlyHisValGlnPhe 437
QY 766 GTGTCTGCTACAGATCCTGTACAGTGTGATGGCAGAAATGGAGTGTGTACAGAGAG 825
Db 438 ValGlySerTyrlLysLeuAlaTySerAsnAspGlyGluHisTrpThrValTyGlnAsp 457
QY 826 CTTGGTGTGGAGCAAGATATTTCAAGGAAACAAAGATATATCACCAGGATGTCGCT 885
Db 458 GluLysGlnArgLysAspLysValPheGlnGlyAlaAsnAspThrHisArgLys 477
QY 886 AATAACTTTTGGCCACCAATTAATGACAGTGTGATGAGTGAATCCTACCCCAATGCG 945
Db 478 AsnValIleAspProIleTyAlaArgHisIleArgIleLeuProTrpSerTrpTyr 497
QY 946 CAGAAATTCATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 981
Db 498 GlyArgIleThrLeuAlaSerGluLeuLeuGlyCys 509

```

```

RESULT 6
US-07-607-538C-4
; Sequence 4, Application US/07607538C
; Patent No. 5455031
; GENERAL INFORMATION:
; APPLICANT: Ceriani Dr., Roberto L.
; APPLICANT: Peterson Dr., Jerry A.
; APPLICANT: Larocca, David J.
; TITLE OF INVENTION: POLYPEPTIDE WITH 46
; TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
; TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
; TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-
; TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
; TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: V. Amzel & Assoc.
; STREET: 2055 No. 5455031th Broadway
; CITY: Walnut Creek
; STATE: California

```



```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Viviana Amzel
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: CRCC-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 943-1931
; TELEFAX: (510) 943-1189
; TELEX: N.A.
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE:
; US-07-607-538C-3

Alignment Scores:
Pred. No.: 4.15e-25 Length: 217
Score: 343.00 Matches: 71
Percent Similarity: 57.23% Conservative: 28
Best Local Similarity: 41.04% Mismatches: 68
Query Match: 9.64% Indels: 6
DB: 1 Gaps: 3

US-10-060-830-1113 (1-1962) x US-07-607-538C-3 (1-217)
QY 472 TTATCTACAGTCTTTTACATTTAAGACAGTGGATGTTATGGAACACTGGGATGGAG 531
Db 48 LeuArgPheGluLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 67
QY 532 TCTGGTGTGATCCGGATCCTCAAAATACAGCATCATCTGTGCTGGAG---TGGACTGAC 588
Db 68 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerTyrLysThrTrpGlyLeu 87
QY 589 CACACAGGGGCAAGACAGATCTGGAACCCAAACCAAGGCTGAAACCACTGCA---645
Db 88 HisLeu-----PheSerTrpAsnProSerTyrAlaArgAsnPheGlySerValGln 104
QY 646 ---CCGCTTGGCTGCTTTTACATTTAAGACAGTGGATGTTATGGAACACTGGGATGGAG 531
Db 105 PheAsnAlaTrpValAlaGlySerTyrGlyAsnAspGlnTrpLeuGlnValAspLeuGly 124
QY 823 GAGCCTGGTGGAGCAAGATATTTCAAGGAAACAAAGATTTATCACCAGGATGTG 882
Db 165 AspProArgThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisLys 184
QY 883 CGTAATAACTTTTGGCCACCAATTTATGACGTTTATTAGATGATCTACCCCAATGG 942
Db 185 LysAsnLeuPheGluThrProIleLeuAlaArgTyrValArgIleLeuProValAlaTrp 204
QY 943 CAGCAGAAATTCCTGAAATGAGGAGCTGCTGGATGT 981
Db 205 HisAsnArgIleAlaLeuArgLeuGluLeuGlyCys 217

RESULT 10
US-08-162-402B-3
; Sequence 3, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LARocca, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-162-402B-3

Alignment Scores:
Pred. No.: 4.15e-25 Length: 217
Score: 343.00 Matches: 71
Percent Similarity: 57.23% Conservative: 28
Best Local Similarity: 41.04% Mismatches: 68
Query Match: 9.64% Indels: 6
DB: 2 Gaps: 3

US-10-060-830-1113 (1-1962) x US-08-162-402B-3 (1-217)
QY 472 TTATCTACAGTCTTTTACATTTAAGACAGTGGATGTTATGGAACACTGGGATGGAG 531
Db 48 LeuArgPheGluLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 67
QY 532 TCTGGTGTGATCCGGATCCTCAAAATACAGCATCATCTGTGCTGGAG---TGGACTGAC 588
Db 68 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerTyrLysThrTrpGlyLeu 87
QY 589 CACACAGGGGCAAGACAGATCTGGAACCCAAACCAAGGCTGAAACCACTGCA---645
Db 88 HisLeu-----PheSerTrpAsnProSerTyrAlaArgAsnPheGlySerValGln 104
QY 646 ---CCGCTTGGCTGCTTTTACATTTAAGACAGTGGATGTTATGGAACACTGGGATGGAG 531
Db 105 PheAsnAlaTrpValAlaGlySerTyrGlyAsnAspGlnTrpLeuGlnValAspLeuGly 124
QY 823 GAGCCTGGTGGAGCAAGATATTTCAAGGAAACAAAGATTTATCACCAGGATGTG 882
Db 165 AspProArgThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisLys 184
QY 883 CGTAATAACTTTTGGCCACCAATTTATGACGTTTATTAGATGATCTACCCCAATGG 942
Db 185 LysAsnLeuPheGluThrProIleLeuAlaArgTyrValArgIleLeuProValAlaTrp 204
QY 943 CAGCAGAAATTCCTGAAATGAGGAGCTGCTGGATGT 981
Db 205 HisAsnArgIleAlaLeuArgLeuGluLeuGlyCys 217
```

QY 823 GAGCCTGGTGGAGCAAGATATTTCAAGCAACAAAGATTATATCACCAGGATGTG 882
Db 165 AspProArgThrGlySerLysIlePheProGlyAsnTrpAsnHisSerHisLys 184
QY 883 CGTAATAACTTTTGGCCACCAATATTGCGCTTTTATTAGAGTGAATCCTACCAATGG 942
Db 185 LysAsnLeuPheGluThrProIleLeuAlaArgTyValArgIleLeuProValAlaTrp 204
QY 943 CAGCAGAAATGCGATGAAATGAGCTGCTCGGATGT 981
Db 205 HisAsnArgIleAlaLeuArgLeuGluLeuGlyCys 217

RESULT 11
US-07-607-538C-2
; Sequence 2, Application US/07607538C
; Patent No. 5455031
; GENERAL INFORMATION:
; APPLICANT: Ceriani Dr., Roberto L.
; APPLICANT: Peterson Dr., Jerry A.
; TITLE OF INVENTION: POLYPEPTIDE WITH 46
; TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
; TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGUES,
; TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-
; TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
; TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: V. Amzel & Assoc.
; STREET: 2055 No. 5455031th Broadway
; CITY: Walnut Creek
; STATE: California
; COUNTRY: USA
; ZIP: 94596
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/607,538C
; FILING DATE: 01-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Viviana Amzel
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: CRFCC-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 943-1931
; TELEFAX: (510) 943-1189
; TELEX: N.A.
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE:
US-07-607-538C-2

Alignment Scores:
Pred. No.: 4,16e-25 Length: 218
Score: 343.00 Matches: 71
Percent Similarity: 57.23% Conservative: 28
Best Local Similarity: 41.04% Mismatches: 68
Query Match: 9.64% Indels: 6
DB: 1 Gaps: 3

US-10-060-830-1113 (1-1962) x US-07-607-538C-2 (1-218)
QY 472 TTATCTACAGCTCTTTTACATTTTAAGCAAGTGGATGTTATGGACACTGGGGATGGAG 531

Db 49 LeuArgPheGluLeuLeuGlyCysGluLeuAsnGlyCysAlaAsnProLeuGlyLeuLys 68
QY 532 TCTGCTGTGATCGGGGATCTCTCAATAACAGCATCATCTGTGCTGGAG---TGGACTGAC 588
Db 69 AsnAsnSerIleProAspLysGlnIleThrAlaSerSerTyLysThrTrpGlyLeu 88
QY 589 CACACAGGGCAAGAGACACAGTTGGAAACCCCAAAAGCCAGCGCTGAAACCACTGGA--- 645
Db 89 HisLeu-----PheSerTrpAsnProSerTyAlaArgLeuAspLysGlnGlyAsn 105
QY 646 ---CCGCTCTGGCTGCTTTGCCACTGATGAATACCACTGGTTCACAAATAGATTGAAT 702
Db 106 PheAsnAlaTrpValAlaGlySerTyGlyAsnAspGlnTrpLeuGlnValAspLeuGly 125
QY 703 AAGGAAAAAGAAATAACAGGCAATTATAACCACTGGATCCACCATGGTGGAGACACAATTAC 762
Db 126 SerSerLysGluValThrGlyIleIleThrGlnGlyAlaArgAsnPheGlySerValGln 145
QY 763 TATGTGCTGCTACAGAACTCTGTACAGTGTATGATGGGAGAAATGAGCTGTGTACAGA 822
Db 146 PheValAlaSerTyLysValAlaTySerAsnAspSerAlaAsnTrpThrGluTyGln 165
QY 823 GAGCCTGTGTGGACCAAGATATTTCAAGCAACAAAGATTATATCACCAGGATGTG 882
Db 166 AspProArgThrGlySerSerLysIlePheProGlyAsnTrpAspAsnHisSerHisLys 185
QY 883 CGTAATAACTTTTGGCCACCAATATTATGCGCTTTTATTAGAGTGAATCCTACCAATGG 942
Db 186 LysAsnLeuPheGluThrProIleLeuAlaArgTyValArgIleLeuProValAlaTrp 205
QY 943 CAGCAGAAATGCGATGAAATGAGCTGCTCGGATGT 981
Db 206 HisAsnArgIleAlaLeuArgLeuGluLeuGlyCys 218

RESULT 12
US-08-162-402B-2
; Sequence 2, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LARocca, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMF) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:

	QY	472	T T A T C A C A A G T C T T T T T A C A T T A T A G A C A A G T G G A T G T A T T G G A C A C A C T G G G G A T G G A G	531
	D b	218	L e u A r g P h e G l u L e u L e u G l y C y s G I u L e u S n A s n P r o L e u G l y L e u L y s	237
	Q Y	532	T C F G G T G T G A T C G C G A T C C T C A A A T A A C A G A T A T C T G T G T G G A G --- - T G G A C T G A C	588
	D b	238 ^a	A s n A s n S e r I l l e P r o A s p l y s G l n I l l e t h r A l a S e r S e r T y r l y s t h r T r p G l y L e u	257
	Q Y	589	C A C A C A G G C C A G A G A C A G T T G G A A A C C C A A A A A A C C A G G C T G A A A A A A C C T G G A --- -	645
	D b	258	H i s L e u -----P h e S e r T r p A s n P r O s e r T y r A l a a r g L e u A s p L y s G l n G l y A s n	274
	Q Y	646	-- --C C G C C T T G G C T C T T T T G C C A C T G A T A C C A C A G T G G T T C A A A T A G A N T R P T G A A T	702
	D b	275	P h e A s n A l a t r p V a l A l a G l y S e r T y r G l y A s n S p l n t r p l e u G l n v a l l a s p L e u G l y	294
	Q Y	703	A A G G A A A A G A A A A T A A C A G C A T T A A A C C A C T G G A T C C A C C A T G T G G A G C A C A A T T A C	762
	D b	295	S e r S e r L y s G l u V a l t h r G l y I l l e t h r G l n G l y A l a a r g A s n P h e G l y S e r V a l G l n	314
	Q Y	763	T A T G T G T C C C T C A G A A T C C T G T C A G T G A T G A T G G G C A G A A A T G G A C A A T G T G T A C A G A	822
	D b	315	P h e V a l A l a S e r T y r L y s V a l A l a T y r S e r A s n A s p S e r A l a A s n t r p T h r G l u T y r G l n	334
	Q Y	823	G A G C C T G C T G C A G C A A G A T A G A T A T T C A A G A A A C A A C A A T T A C C A C A G G A T G T G	882
	D b	335	A s p P r o A r g T h r G l y S e r S e r L y s I l l e P h e P r o G l y A s n T r p A s n H i s S e r H i s L y s	354
	Q Y	883	C G T A A T A A C T T T T T G C C A C A A T A T T G C A C G T T T T A T T A G A C T G A A T C C T A C C C A A T G G	942
	D b	355	L y s A s n L e u P h e G l u T h r P r o l l e L e u A l a a r g T y r V a l a r g l l e L e u P r o V a l A l a t r p	374
	Q Y	943	C A G C A A A A T T G C A T G A A A A T G G A G C T C T C G G A T G T	981

Best Local Similarity: 42.59% Mismatches: 61
Query Match: 9.58% Indels: 6
DB: 2 Gaps: 3

US-10-060-830-1113 (1-1962) x US-08-162-402B-12 (1-159)

QY 505 GGATGTTTGAACACTGGGATGCTGCTGGTGGCGGATCCTCAATAACACGA 564
|||||
Db 1 GlyCysAlaAsnProLeuGlyLeuLysAsnSerIleProAspLysGlnIleThrAla 20
|||||
QY 565 TCATCTGTGCTGGAG---TGGACTGACCACACAGGCGCAAGAGAACAGTTGGAACCCAAA 621
|||||
Db 21 SerSerSerTyrLysThrTrpGlyLeuHisLeu-----PheSerTrpAsnProSer 37
|||||
QY 622 AAAGCCAGGCTGAAAAACCTGGA-----CCGCTTGGGCTGCTTTTGGCCACTGATGAA 675
|||||
Db 38 TyrAlaArgLeuAspLysGlnGlyAsnPheAsnAlaTrpValAlaGlySerTyrGlyAsn 57
|||||
QY 676 TACCAGTGGTTACAATAGATTGTAATAGGAAAGAAATAACAGGCATTATACCACT 735
|||||
Db 58 AspGlnTrpLeuGlnValAspLeuGlySerLysGluValThrGlyIleThrGln 77
|||||
QY 736 GGATCCACCATGGTGGAGCACAAATTACTATGTCTGCTACAGAACTCTGTACAGTGAT 795
|||||
Db 78 GlyAlaArgAsnPheGlySerValGlnPheValAlaSerTyrLysValAlaTyrSerAsn 97
|||||
QY 796 GATGGCAGAAATGGACTGTGTACAGAGAGCGCTGTGTGGAGCAAGATAAGATATTCAA 855
|||||
Db 98 AspSerAlaAsnTrpThrGluTyrGlnAspProArgThrGlySerSerLysIlePhePro 117
|||||
QY 856 GGAACAAGATTATCACCAGGATGTGGTAATACTTTTGGCCACCAATTATTCACGT 915
|||||
Db 118 GlyAsnTrpAspAsnHisSerHisLysLysAsnLeuPheGluThrProIleLeuAlaArg 137
|||||
QY 916 TTTATTAGAGTCAATCCTACCAATGGCAGCAGAGAAATGGCAATGGAGCTGCTC 975
|||||
Db 138 TyrValArgIleLeuProValAlaTrpHisAsnArgIleAlaLeuArgLeuGluLeu 157
|||||
QY 976 GGATGT 981
|||||
Db 158 GlyCys 159

Search completed: January 21, 2003, 09:51:29
Job time : 21.1166 secs

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 192247)
Birken B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,
Baldwin J., Barna N., Beckerly R., Boguslavsky L., Boukhgalter B.,
Brown A., Castie A., Collangelo M., Collins S., Collymore A.,
Cooke P., DeArellano K., Dewar K., Domino M., Donelan L., Doyle M.,
Ferreira P., FitzHugh W., Forrest C., Funke R., Gage D.,
Galagan J., Gardyna S., Grant G., Hagos B., Heaford A., Horton L.,
Howland J.C., Johnson R., Jones C., Kann L., Karatas A., Klein J.,
Lehoczyk J., Lieur C., Locke K., Macdonald P., Marquis N.,
McEwan P., McGurk A., McKernan K., McLaughlin J., Meldrim J.,
Morrow J., Naylor J., Norman C.H., O'Connor T., O'Donnell P.,
Peterson K., Pollara V., Riley R., Roy A., Santos R., Severy P.,
Stange-Thomann N., Stojanovic N., Subramanian A., Talamas J.,
Tesfaye S., Tirrell A., Vassiliev H., Vo A., Wheeler J., Wu X.,
Wyman D., Ye W.J., Zimmer A. and Zody M.

Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6692347.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1852
Center clone name: 319_J_24
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182870 bases at least Q40
Consensus quality: 187032 bases at least Q30
Consensus quality: 188925 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 191247; sum-of-contents
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-contents

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 472: contig of 472 bp in length
* 473 572: gap of 100 bp
* 573 3945: contig of 3373 bp in length
* 3946 4045: gap of 100 bp
* 4046 11056: contig of 7011 bp in length
* 11057 11156: gap of 100 bp
* 11157 17728: contig of 6572 bp in length
* 17729 17828: gap of 100 bp
* 17829 30556: contig of 12728 bp in length
* 30557 30656: gap of 100 bp
* 30657 49294: contig of 18638 bp in length
* 49295 49394: gap of 100 bp
* 49395 64509: contig of 15115 bp in length
* 64510 64609: gap of 100 bp
* 64610 83042: contig of 18433 bp in length
* 83043 83142: gap of 100 bp
* 83143 112373: contig of 29231 bp in length
* 112374 112473: gap of 100 bp
* 112474 145663: contig of 33190 bp in length
* 145664 145763: gap of 100 bp
* 145764 192247: contig of 46484 bp in length.

FEATURES

Location/Qualifiers

Source

1. 192247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-319J24"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1. 472
/note="assembly_fragment"
clone_end:SP6
vector_side:right
573. 3945
/note="assembly_fragment"
4046. 11056
/note="assembly_fragment"
11157. 17728
/note="assembly_fragment"
17829. 30556
/note="assembly_fragment"
30657. 49294
/note="assembly_fragment"
49395. 64509
/note="assembly_fragment"
64610. 83042
/note="assembly_fragment"
clone_end:T7
vector_side:left
83143. 112373
/note="assembly_fragment"
112474. 145663
/note="assembly_fragment"
145764. 192247
/note="assembly_fragment"
BASE COUNT 58524 a 38157 c 36840 g 57721 t 1005 others
ORIGIN

Alignment Scores:
Pred. No.: 9.82e-52 Length: 192247
Score: 922.50 Matches: 201
Percent Similarity: 44.87% Conservative: 0
Best Local Similarity: 44.87% Mismatches: 1
Query Match: 26.59% Indels: 247
DB: 2 Gaps: 1

US-10-060-830-1114 (1-653) x AC013497 (1-192247)

QY 452 GYTrpTlpGlyGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThr 471
|||||
Db 164102 GGTGTGGTGAAGAAATGAGCAGCTTCTTCCTGCAAAAGCAGTGGACCAATGAGGAACC 164161
|||||

QY 472 ProValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrVal 491
|||||
Db 164162 CCAGTTCGCTATAGCAGCAGCGAAGTTAATCACCTGAGTCCAAGAGAAGTCACCACAGTG 164221
|||||

QY 492 LeuGlnAlaAspSerAla----- 497
|||||

Db 164222 CTGCAGGCTGACTCTGC-AGGTAACATATGTTCGACGCTTCTGTGTACCAGCAGAGGAGA 164280
QY 497 ----- 497

Db 164281 AACTGCTTAGGCTGTGTATAAAGTGTGTTGGGATTTTACAGTTCTTGTATCCCTTCATGT 164340
QY 497 ----- 497

Db 164341 TTAAGAAATGAATGTTTCTCTAAAGCTAGAACCCACATTTTAAAGAGTGACTCTTCAGAG 164400
QY 497 ----- 497

Db 164401 AGTCATTTCTGTATCTTGGATATTTCTGTCTGTAAAGTTGGACATGATTTCTTAAAGACT 164460
QY 497 ----- 497

Db 164461 CTGGGCATTTATGTATGTTAACTGAGGCTTGTGGGGAGACCCGAGAGTGTGTGAGAGGAGC 164520

QY	497	-----	497	Homo sapiens
Db	164521	GGGAAGACAGAGAAGCGCTTACAGTGTTATCTCTTAGCCTAGTACATCTTTATTTA	164580	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
QY	497	-----	497	REFERENCE
QY	497	-----	497	1 (bases 1 to 193623)
Db	164581	AACATTCTCTAATATTGAAAAAATCTCTGAATTAATTTTCAAGATAATATTTATTC	164640	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyne,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S., Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and Gibbs,R.
QY	497	-----	497	Direct Submission
QY	497	-----	497	Unpublished
QY	497	-----	497	2 (bases 1 to 193623)
QY	497	-----	497	Worley,K.C.
QY	497	-----	497	Direct Submission
QY	497	-----	497	Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
QY	497	-----	497	3 (bases 1 to 193623)
QY	497	-----	497	Worley,K.C.
QY	497	-----	497	Direct Submission
QY	497	-----	497	Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
QY	497	-----	497	4 (bases 1 to 193623)
QY	497	-----	497	Worley,K.C.
QY	497	-----	497	Direct Submission
QY	497	-----	497	Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
QY	497	-----	497	On May 25, 2002 this sequence version replaced gi:20429236.
QY	497	-----	497	INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu
QY	497	-----	497	CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of
QY	497	-----	497	COMMENT
QY	497	-----	497	RESULT 10
QY	497	-----	497	AC106728
QY	497	-----	497	LOCUS
QY	497	-----	497	DEFINITION
QY	497	-----	497	AC106728 Homo sapiens 3 BAC RP11-319J24 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
QY	497	-----	497	AC106728
QY	497	-----	497	AC106728.6 GI:21206081
QY	497	-----	497	KEYWORDS
QY	497	-----	497	HTG.

the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT:

FEATURES	Location/Qualifiers
source	1. 193623
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clone="RP11-319J24"
repeat_region	818..1088
	/rpt_family="L2"
	complement(3791..3934)
repeat_region	/rpt_family="MIR"
	complement(4203..4581)
repeat_region	/rpt_family="L2"
	4767..4790
	/rpt_family="(TC)n"
	4791..4851
repeat_region	/rpt_family="(TC)n"
	4854..4892
repeat_region	/rpt_family="(CA)n"
	4907..5066
repeat_region	/rpt_family="L2"
	5333..5386
	/rpt_family="GA-rich"
repeat_region	complement(5512..5945)
	/rpt_family="MLTIC"
repeat_region	complement(6187..6624)
	/rpt_family="LTR37B"
repeat_region	complement(6858..7219)
	/rpt_family="L2"
STS	6880..7000
	/standard_name="73424"
repeat_region	complement(7409..7632)
	/rpt_family="MIR"
repeat_region	7862..8369
	/rpt_family="L1MC1"
repeat_region	8417..8524
	/rpt_family="L1M4"
repeat_region	complement(10026..10215)
	/rpt_family="MER20"
repeat_region	complement(10485..10549)
	/rpt_family="MER112"
repeat_region	12814..13590

repeat_region	/rpt_family="L1PA7"
	13917..14135
repeat_region	/rpt_family="L1M5A"
	complement(14753..15070)
repeat_region	/rpt_family="Alusq"
	15152..15366
repeat_region	/rpt_family="MIR"
	15395..15562
repeat_region	/rpt_family="MER5B"
	16109..16388
repeat_region	/rpt_family="L1MC5"
	16389..16642
repeat_region	/rpt_family="AluY"
	16643..16758
repeat_region	/rpt_family="L1MC5"
	16826..16898
repeat_region	/rpt_family="L1MC5"
	complement(17591..18655)
repeat_region	/rpt_family="L1MC4"
	complement(19191..19653)
repeat_region	/rpt_family="L1MD3"
	complement(19662..19780)
repeat_region	/rpt_family="FLAM-A"
	complement(19781..21049)
repeat_region	/rpt_family="L1MD3"
	complement(21069..21455)
repeat_region	/rpt_family="MLT1B"
	complement(21479..22318)
repeat_region	/rpt_family="L1MD3"
	complement(22447..22681)
repeat_region	/rpt_family="L1MD3"
	complement(23470..23853)
repeat_region	/rpt_family="MSTD"
	23852..23925
repeat_region	/rpt_family="U3"
	23936..25203
repeat_region	/rpt_family="L1PB3"
	24844..24975
STS	/standard_name="6234"
	25219..25388
repeat_region	/rpt_family="(TTTC)n"
	complement(25428..25548)
repeat_region	/rpt_family="Aluub"
	25555..25950
repeat_region	/rpt_family="L1PB3"
	25951..26456
repeat_region	/rpt_family="MER75"
	26457..27565

Alignment Scores:

Pred. No.:	9.91e-52	Length:	193623
Score:	922.50	Matches:	201
Percent Similarity:	44.87%	Conservative:	0
Best Local Similarity:	44.87%	Mismatches:	1
Query Match:	26.59%	Indels:	247
DB:	9	Gaps:	1

US-10-060-830-1114 (1-653) x AC106728 (1-193623)

OY	452	GLYTPPLYSGLYMETLGSGLNPhenauProAlaYsAlaValAspHisGluGluThr	471
DB	95315	GGTTGCTGGAAAGAAATAGACAGTTTCTTCGAAAGAGGAGGACATGAGGAAAC	95374
OY	472	ProValArgTYrSerSerSeiGluValAsnHisLeuSerProArgGluValThrThyAl	491
DB	95375	CCAGTTCCTATAGCAGCAGGAGGAGTATATCCTGATCCAAAGAGAGTACCAACATG	95434
OY	492	LeuGlnAlaAspSerAla	497
DB	95435	CTGCAGGCTGACTCTGC-AGGTAACATATGTTGACACCTTCTGTCACAGGAGGAGGA	95493
OY	497	-----	497

```
Db 95494 AACTGCTTAGGCTTGCTATAAAGTCTTTGGGATTACAGTTCTTTGATCCCTTTTCATGT 95553
QY 497 ----- 497
Db 95554 TTAAGAATGAATGTTTCTTAAGAGTAGAACCACTTTTTTAAAGGTGACTCTTCAGAG 95613
QY 497 ----- 497
Db 95614 AGTCATTTCTGTATCTGGGATATCTCTGTAAGTGGGACATTGATTTCTAAAGACT 95673
QY 497 ----- 497
Db 95674 CTGGCATTATGTATGTTAACTGAGGCTTCTGGGGAGACCCGAGATGTTGAGAGGAGC 95733
QY 497 ----- 497
Db 95734 GGGAAAGACAGAAAGCCCTTACAGTCTTTATCTCTTAGCCCTAGTACATACATATTTTA 95793
QY 497 ----- 497
Db 95794 AACATTTCTTAATTTGAAAAAATCTCGAAATTTATTTTCAAGATTAATAATTTATTC 95853
QY 497 ----- 497
Db 95854 CAGTAACGTAATGTTTCTTTTCCAAATGAATATTTCTGAAAAATATTTCAAGAAAGGAATAT 95913
QY 497 ----- 497
Db 95914 TCAGTAAGAATAATGTTGCTGTGTAATCTTAAGAACATATGTGAAAATTTCTATATGCAA 95973
QY 497 ----- 497
Db 95974 AATTACATGATGAGACTTCTGCTATATTTACTATGAGAAGTACTAGACAGTATGTT 96033
QY 497 ----- 497
Db 96034 TAATTTTAAATTTCAAAGTCTTACAGATTAAGTATTGATGGCAACAAAGTATCCTAAA 96093
QY 497 ----- 497
Db 96094 GAATGTTTGTCTTTAGTATAATCTGCAGCTCCACCCCTAAATTAATGCAATCTCTTTACTA 96153
QY 498 -----GlutYrAlaGlnProLeuValGly 505
Db 96154 GTAAATGACTTTTCTGTGTTTGTGTATATCCACAGATGATGCTCAGCCACTGCTAGGA 96213
QY GlyTleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGluAla 525
Db 96214 GGAATTTGTTGTACACTTTCATCAAGATCTACCTTTAAACCAAGAGAAAGGAAGCA 96273
QY GlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHisAlaTyrAla 545
Db 96274 GGCTATGAGACCTAGATCTTACAACTCAGCAGGGCAGGAAGTTTATCATGCTCTATGCT 96333
QY 546 GluProLeuProIleThrGlyProGluTyrAlaThrProIleMetAspMetSerGly 565
Db 96334 GAACACATCCCAATTTACGGGCTGAGTATGCAACCCCAATCATCATGACATGTCAGGG 96393
QY HisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGln 585
Db 96394 CACCCCACAACTTCAGTTGGTCAGCCCTCCACATCCACTTCAAGGCTACGGGGAACCAA 96453
QY ProProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSerCysSerSer 605
Db 96454 CCTCCCCCACTAGTGGGAACCTTACAATACATCTCTCCAGGACTGACAGCTGCTCCCTCA 96513
QY 606 AlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuProAlaProAspGlu 625
Db 96514 GCCCAGGGCCAGTATGATACCCGGAAGCTGGGAAGCCAGGCTTACCTGCCCCCAGACGAA 96573
QY 626 LeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGlyGlu 645
Db 96574 TTGGTGTACCGGTGCCACAGACGACACAAAGAAGTATCAGGAGCAGGAAGGATGGGAA 96633
```

```
QY 646 CysaspValPheLysGluIleLeu 653
Db 96634 TGTGATCTTTTAAAGAAATCCTT 96657
```

RESULT 11

AK095973

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

CDS

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

AK095973 2547 bp mRNA linear PRI 15-JUL-2002
Homo sapiens cDNA FLJ38654 fis, clone HHDPC2009178, weakly similar
to Coagulation factor V.

AK095973 1 GI:21755342
oligo capping: fis (full insert sequence).
Homo sapiens dermal papilla cells (HDPC) cDNA to mRNA,
clone lib: HHDPC2 clone: HHDPC2009178.

Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1
Nishi, T., Ota, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,
Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K.,
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,
Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished

2 (bases 1 to 2547)
Isogai, T. and Yamamoto, J.
Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team): 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end, one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

Location/Qualifiers
1. 2547

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HHDPC2009178"
/cell_type="dermal papilla cells (HDPC)"
/clone_lib="HHDPC2"
/note="cloning vector: pME18SFL3-primary culture, dermal
papilla cells"
49..1623
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC04663.1"
/db_xref="GI:21755343"
/translation="MVPARGGGALARAAGRLALLLAVSAPLRLQAELEGDCGCHL
VTYODSGTMTSKDYPGTYNHTVCEKTTVPKRLILRGLDIESOTCASDYLFT
SSDOQGYPCSMVTVPKELLNLTSEVTFEESGSHISGRGFLTYASDHPDLPLCLE
RASHLYKTEYSKFCPAGCDVAGDISGNVDGRTSLCKNAIHAGIIRASSSQSV
VLQRGIGSYEGILANGVLSRDSLSKDLFTSNGCSRSLSLIDGERKKKIGITGTSQ
NESDQVHWSFGQARLQDQSWASGSDSNHNNHKEWLEIDIGERKKKIGITGTSQ
SNFNYKVSFVNNKNSKWTYKGIYNNNEKVFQNSRFPQNNPPIPIVARY
RVPTQTHORIALKVELICCOITGNDLSLWVKTSQTSVSTPKKEDETTRIPSEET
STGNTITVAIPVLVLVVFAGMGIFRAFRKRERKVRMDQDLRQKTVGSRNLNPLP
DISQLSPSAMIIMRRR"

646 a 632 c 695 g 574 t
BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 1.88e-50 Length: 2547
Score: 870.00 Matches: 220
Percent Similarity: 50.08% Conservative: 107

Best Local Similarity: 33.69% Mismatches: 229
 Query Match: 25.07% Indels: 98
 DB: 9 Gaps: 17

US-10-060-830-1114 (1-653) x AK095973 (1-2547)

QY 13 LeuLeuLeuAspAlaGlyAlaGlnGlnGlyLysCysGlyLeuGlyLeuGlnMet 32
 DB 343 CTTCTCTCACCAGCTCTTGCAGATCAATATGCTCCACTACTGTGGA---AGTATGACTGTT 399
 QY 33 AsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHis 52
 DB 400 CCCAAGAACTCTTGTGTGAACACAGTGAAGTAACCGTCGCTTGTGAGAGTGGATCCAC 459
 QY 53 ValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLysLeuIleThr 72
 DB 460 ATTTCGCGCGGGGTTTTTGTGCTACCTATGCGCAGCAGCACCACCATCAAGATTAATAACA 519
 QY 73 CysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGly 92
 DB 520 TGTGTTGGAACGAGCTAGCCATTAATTTCAAGACAGATAACAGCAAAATCTGCCAGCTGTT 579
 QY 93 CysLeuLeuProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerSer 112
 DB 580 TGTAGAGACGTACAGAGACATTTCTGGAATATGCTAGATGATATAGAGATACCTCT 639
 QY 113 ProLeuCysMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIle 132
 DB 640 TTATTGTGCAAGCTGCCATCCATGAGGAATAATTGCTGATCAACTAGGTGCCAGATC 699
 QY 133 SerValValIleSerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThr 152
 DB 700 AGTGTGCTTCAGCGCAAGGGATCACTCGATATGAAGGATTCGTGCGCAATGTTGTTCT 759
 QY 153 SerValValGlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGly 172
 DB 760 TCAGGGAGGTTCCTGTGTCAGACAGCGATTCTGTTTACCTCCCAATGTTGCAGCAGA 819
 QY 173 ThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeu 192
 DB 820 TCCTTGAGTTTGAA-----CCTGACGGGCAAAATCAGAGCTTCTCTCATGG 867
 QY 193 GluTrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLys 212
 DB 868 CAGTCGGTCAATGAGATGGAGACCAAGTTCAGTGTCTCTCGGCCCAAGCCGACCTCAG 927
 QY 213 LysProGlyProProTrpAlaAlaPheAlaThrAspGlu-----TyrGlnTrp 228
 DB 928 GACCAAGGCCCATCATGGGCTTCGGCGCAGTAGCAACACCAACACCAAGAGATGG 987
 QY 229 LeuGlnIleAspLeuAsnLysGlyLysLysIleThrGlyIleThrThrGlySerThr 248
 DB 988 CTGGAGATCGATTGGGGGAGAAAAAGAAAAAATAACAGGAATTAGGACACACAGGATCA 1047
 QY 249 MetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGln 268
 DB 1048 CAGTCGAACCTCAACTTTTATGTAAGAGTTTGTGATGAACCTCAAAAACAATAATCT 1107
 QY 269 LysTrpThrValTyrArgGluProGlyValGluAspLysIlePheGlnGlyAsnLys 288
 DB 1108 AGTGGNAGACCTATAAGGAATGTGAATAAAGAAAGAGTTGTTTCAAGGTAACTCT 1167
 QY 289 AspTyrHisGlnAspValArgAsnAsnPheLeuProProIleAlaArgPheIleArg 308
 DB 1168 AACTTTTGGGACCCAGTGCAAAACAATTTTCATCCCTCCCATCGTGGCCAGATATGTGCGG 1227
 QY 309 ValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGln 328
 DB 1228 GTTGTCCCCAGCATGGCACCAGGATAGCCTTCAAGGTGGAGCTCAATGTTGTTGCCAG 1287
 QY 329 PheIleProLysGlyArgProProLysLeuThrGlnProProProProArgAsnSerAsn 348
 DB 1288 -----ATTACAAA-----GGTAAT 1302

QY 349 AspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPheThr 368
 DB 1303 GAT----- 1305
 QY 369 GlnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrAla 388
 DB 1306 -----TCATTGTTGTGGCGCAGACAGCAAGTCAAGCACCAGGTGTT 1344
 QY 389 SerProAspIleArgAsnThrThrValThrProAsnValThrLysAspValAlaLeuAla 408
 DB 1345 TCAACTAAGAAAGAAAGATGAGACAATCACAAGGCCCATCCCTCGGAGAAACATCCACA 1404
 QY 409 AlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeuValCys 428
 DB 1405 GGAATAAACATTACAACGGTGGCTATTCCATTGCTGCTCTCTGTTGCTCTGTTG 1463
 QY 429 AlaTrpHisTrpArgAsnArg-----LysLysLysLysThrGluGlyThrTyr 443
 DB 1464 -----TGAATGGGATCTTTTGCAGCCTTTAGAAAAAAGAAAGAGGAGTCCGTAT 1514
 QY 444 AspLeuProTyrTrpAspArgAlaGlyTrpLysGlyMetLysGlnPheLeuProAla 463
 DB 1515 GGATCACAGAGGCTCAGAAACAGACTGTTGGAGCAGATTAAATATCCCTTT----- 1568
 QY 464 LysAlaValAspHisGluGluThrProValArgTyrSerSerSerGlu-----ValAsn 481
 DB 1569 -----GCCAGACATCAGTCAGCTGAGTTTACCATCAGC 1601
 QY 482 HisLeuSerProArgGluValThr-----ThrValLeuGlnAlaAspSerAlaGlu 498
 DB 1602 TATGATAATGAGAAGGATGATGACACAAAGTTAGATCTCATCAAGTGATATGGCAGAT 1661
 QY 499 TyrAlaGlnProLeuValGlyIleValGlyThrLeuHisGlnArgSerThrPheLys 518
 DB 1662 TACCAGCAGCCCTCATGATTTGGCACCAGGACAGTACGAGGAGGGCTCCACCTTCGG 1721
 QY 519 Pro-----GluGluGlyLysGluAlaGlyTyrAlaAspLeuAspPro-----Tyr 533
 DB 1722 CCCATGGACACGGATGCGGAGGCGAGG-----GTGAGCACCAGTCCGCGGCGGCACAT 1778
 QY 534 AsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuIleThrGlyPro 553
 DB 1779 GACTGCCCGCAGCG 1838
 QY 554 GluTyrAlaThrProIleIleMetAspMetSerGlyHisProThrThrSerValGlyGln 573
 DB 1839 GAGTACGCCAGCCCATCGTG-----GAGCGGACAGTCTGCGCGCGCCACACAGTTC 1889
 QY 574 ProSerThrSerThrPheLysAlaThrGlyAsnGlnPro----- 586
 DB 1890 TCTCGCAGAGCGGCTACCGGTCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1949
 QY 587 -----ProProLeu-----ValGlyThrTyrAsnThrLeuLeuSerArgThr 600
 DB 1950 TCGGGCGGTTCTCCCGGTAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2009
 QY 601 AspSerCysSerSerAlaGlnAlaGlnThrAspThrProLysAla-GlyLysProGlyLe 620
 DB 2010 CACAGCGCACAGCTCGCGCAGGGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2069
 QY 620 uProAlaProAspGluLeuValTyrGlnValProGln 632
 DB 2070 ACCGAAAGCGGACACCCCTGACTCTCAGAAAGCGCCCAA 2106

RESULT 12
 AC126975/c 203102 bp DNA linear HTG 11-JUL-2002
 LOCUS Rattus norvegicus clone CH230-234N3, *** SEQUENCING IN PROGRESS
 DEFINITION *** 74 unordered pieces.
 ACCESSION AC126975 GI:21731383
 VERSION AC126975.1
 KEYWORDS HTG; HTGS_PHASE1.

SOURCE
ORGANISM
REFERENCE
AUTHORS

Norway rat.

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 203102)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbala,J., Benton,J., Bimaga,K., Blankenburg,K., Bonin,D.,
Bouch,J., Bowle,S., Brleva,M., Brown,M., Brown,M., Bryant,N.P.,
Buhay,A., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Hollway,C., Hollins,B.,
Homs,J., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivel,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovick,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Louised,H.,
Locado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mashiney,E., McLeod,M.P., Mendor,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,J., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogub,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherrer,S., Scott,G., Shen,H., Shooshari,N., Slisson,I.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,Y., Villalton,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watling,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Wielstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 203102)

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL

Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: CH230-234N3

Center clone name: CH230-234N3

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 145293 bases at least Q40

Consensus quality: 154399 bases at least Q30

Consensus quality: 161802 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently

* consists of 74 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N's, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	1310	contig of 1310 bp in length
1311	1410	gap of unknown length
1411	2497	contig of 1087 bp in length
2498	2597	gap of unknown length
2598	3718	contig of 1121 bp in length
3719	3818	gap of unknown length
3819	4974	contig of 1156 bp in length
4975	5074	gap of unknown length
5075	6118	contig of 1044 bp in length
6119	6218	gap of unknown length
6219	7276	contig of 1058 bp in length
7277	7376	gap of unknown length
7377	8394	contig of 1018 bp in length
8395	8494	gap of unknown length
8495	9899	contig of 1405 bp in length
9900	9999	gap of unknown length
10000	11010	contig of 1011 bp in length
11011	11110	gap of unknown length
11111	12429	contig of 1319 bp in length
12430	12529	gap of unknown length
12530	14233	contig of 1704 bp in length
14234	14333	gap of unknown length
14334	15599	contig of 1266 bp in length
15600	15699	gap of unknown length
15700	16726	contig of 1027 bp in length
16727	16826	gap of unknown length
16827	18918	contig of 2092 bp in length
18919	19018	gap of unknown length
19019	20837	contig of 1819 bp in length
20838	20937	gap of unknown length
20938	23036	contig of 2099 bp in length
23037	23136	gap of unknown length
23137	24890	contig of 1734 bp in length
24891	24990	gap of unknown length
24991	26683	contig of 1693 bp in length
26684	26783	gap of unknown length
26784	28087	contig of 1304 bp in length
28088	28187	gap of unknown length
28188	29847	contig of 1660 bp in length
29848	29947	gap of unknown length
29948	31165	contig of 1218 bp in length
31166	31265	gap of unknown length
31266	32327	contig of 1062 bp in length
32328	32427	gap of unknown length
32428	33354	contig of 1127 bp in length
33355	33654	gap of unknown length
33655	34694	contig of 1040 bp in length
34695	34794	gap of unknown length
34795	36507	contig of 1713 bp in length
36508	36607	gap of unknown length
36608	38322	contig of 1715 bp in length
38323	38422	gap of unknown length
38423	40396	contig of 1974 bp in length
40397	40496	gap of unknown length
40497	42695	contig of 2199 bp in length
42696	42795	gap of unknown length
42796	43853	contig of 1058 bp in length
43854	43953	gap of unknown length
43954	45008	contig of 1055 bp in length
45009	45108	gap of unknown length
45109	47026	contig of 1918 bp in length
47027	47126	gap of unknown length
47127	49195	contig of 2069 bp in length
49196	49295	gap of unknown length
49296	51748	contig of 2453 bp in length
51749	51848	gap of unknown length

Seq ID	Seq Name	Length	Score	Percent Similarity	Best Local Similarity	Query Match	DB
51849	54202: contig of 2354 bp in length	6.6e-47	852.50	58.36%	54.41%	24.57%	2
54203	54302: gap of unknown length	179	58.36%	54.41%	24.57%	2	112974
54303	56606: contig of 2304 bp in length	13	58.36%	54.41%	24.57%	2	112974
56607	56706: gap of unknown length	35	58.36%	54.41%	24.57%	2	112974
58300	58300: contig of 1594 bp in length	103	58.36%	54.41%	24.57%	2	112974
58301	58400: gap of unknown length	3	58.36%	54.41%	24.57%	2	112974
58401	59928: contig of 1528 bp in length	179	58.36%	54.41%	24.57%	2	112974
59929	60028: gap of unknown length	13	58.36%	54.41%	24.57%	2	112974
60029	62227: contig of 2199 bp in length	35	58.36%	54.41%	24.57%	2	112974
62228	62327: gap of unknown length	103	58.36%	54.41%	24.57%	2	112974
62328	64445: contig of 2118 bp in length	179	58.36%	54.41%	24.57%	2	112974
64446	64545: gap of unknown length	3	58.36%	54.41%	24.57%	2	112974
64546	66642: contig of 2097 bp in length	179	58.36%	54.41%	24.57%	2	112974
66643	66742: gap of unknown length	13	58.36%	54.41%	24.57%	2	112974
66743	68545: contig of 1803 bp in length	35	58.36%	54.41%	24.57%	2	112974
68546	68645: gap of unknown length	103	58.36%	54.41%	24.57%	2	112974
68646	70370: contig of 1725 bp in length	179	58.36%	54.41%	24.57%	2	112974
70371	70470: gap of unknown length	3	58.36%	54.41%	24.57%	2	112974
70471	72796: contig of 2326 bp in length	179	58.36%	54.41%	24.57%	2	112974
72797	72896: gap of unknown length	13	58.36%	54.41%	24.57%	2	112974
72897	75152: contig of 2256 bp in length	35	58.36%	54.41%	24.57%	2	112974
75153	75252: gap of unknown length	103	58.36%	54.41%	24.57%	2	112974
75253	77649: contig of 2397 bp in length	179	58.36%	54.41%	24.57%	2	112974
77650	77749: gap of unknown length	3	58.36%	54.41%	24.57%	2	112974
77750	80695: contig of 2946 bp in length	179	58.36%	54.41%	24.57%	2	112974
80696	80795: gap of unknown length	13	58.36%	54.41%	24.57%	2	112974
80796	83797: contig of 3002 bp in length	35	58.36%	54.41%	24.57%	2	112974
83798	83897: gap of unknown length	103	58.36%	54.41%	24.57%	2	112974
83898	86763: contig of 2866 bp in length	179	58.36%	54.41%	24.57%	2	112974
86764	86863: gap of unknown length	3	58.36%	54.41%	24.57%	2	112974
86864	89822: contig of 3029 bp in length	179	58.36%	54.41%	24.57%	2	112974
89893	89992: gap of unknown length	13	58.36%	54.41%	24.57%	2	112974
89993	92105: contig of 2113 bp in length	35	58.36%	54.41%	24.57%	2	112974
92106	92205: gap of unknown length	103	58.36%	54.41%	24.57%	2	112974
92206	94330: contig of 2125 bp in length	179	58.36%	54.41%	24.57%	2	112974
94331	94430: gap of unknown length	3	58.36%	54.41%	24.57%	2	112974
94431	96835: contig of 2405 bp in length	179	58.36%	54.41%	24.57%	2	112974
96836	96935: gap of unknown length	13	58.36%	54.41%	24.57%	2	112974
96936	99553: contig of 2618 bp in length	35	58.36%	54.41%	24.57%	2	112974
99554	99653: gap of unknown length	103	58.36%	54.41%	24.57%	2	112974
99654	102601: contig of 2948 bp in length	179	58.36%	54.41%	24.57%	2	11

Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczyk, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 146515)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczyk, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 12, 2000 this sequence version replaced gi:7230271.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7942

Center clone name: 322_A_20

* NOTE: This record contains 164 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 836: contig of 836 bp in length

837 936: gap of 100 bp

937 1748: contig of 812 bp in length
 1749 1848: gap of 100 bp
 1849 2680: contig of 832 bp in length
 2681 2780: gap of 100 bp
 2781 3534: contig of 754 bp in length
 3535 3634: gap of 100 bp
 3635 4383: contig of 749 bp in length
 4384 4483: gap of 100 bp
 4484 5261: contig of 778 bp in length
 5262 5361: gap of 100 bp
 5362 6136: contig of 775 bp in length
 6137 6236: gap of 100 bp
 6237 7045: contig of 809 bp in length
 7046 7145: gap of 100 bp
 7146 7953: contig of 808 bp in length
 7954 8053: gap of 100 bp
 8054 8857: contig of 804 bp in length
 8858 8957: gap of 100 bp
 8958 9784: contig of 827 bp in length
 9785 9884: gap of 100 bp
 9885 10718: contig of 834 bp in length
 10719 10818: gap of 100 bp
 10819 11640: contig of 822 bp in length
 11641 11740: gap of 100 bp
 11741 12534: contig of 794 bp in length
 12535 12634: gap of 100 bp
 12635 13407: contig of 773 bp in length
 13408 13507: gap of 100 bp
 13508 14276: contig of 769 bp in length
 14277 14376: gap of 100 bp
 14378 15185: contig of 809 bp in length
 15186 15285: gap of 100 bp
 15286 16086: contig of 801 bp in length
 16087 16186: gap of 100 bp
 16187 16986: contig of 800 bp in length
 16987 17086: gap of 100 bp
 17088 17913: contig of 827 bp in length
 17914 18013: gap of 100 bp
 18014 18841: contig of 828 bp in length
 18842 18941: gap of 100 bp
 18942 19770: contig of 829 bp in length
 19771 19870: gap of 100 bp
 19871 20689: contig of 819 bp in length
 20690 20789: gap of 100 bp
 20790 21575: contig of 786 bp in length
 21576 21675: gap of 100 bp
 21676 22423: contig of 748 bp in length
 22424 22523: gap of 100 bp
 22524 23322: contig of 799 bp in length
 23323 23422: gap of 100 bp
 23423 24239: contig of 817 bp in length
 24240 24339: gap of 100 bp
 24340 25199: contig of 860 bp in length
 25200 25299: gap of 100 bp
 25300 26131: contig of 832 bp in length
 26132 26231: gap of 100 bp
 26232 27103: contig of 872 bp in length
 27104 27203: gap of 100 bp
 27204 28060: contig of 857 bp in length
 28061 28160: gap of 100 bp
 28161 29013: contig of 853 bp in length
 29014 29113: gap of 100 bp
 29114 29944: contig of 831 bp in length
 29945 30044: gap of 100 bp
 30045 30746: contig of 702 bp in length
 30747 30846: gap of 100 bp
 30847 31652: contig of 806 bp in length
 31653 31752: gap of 100 bp
 31753 32559: contig of 807 bp in length
 32560 32659: gap of 100 bp
 32660 33470: contig of 811 bp in length
 33471 33570: gap of 100 bp
 33571 34386: contig of 816 bp in length

```

* 34387 34486: gap of 100 bp
* 34487 35290: contig of 804 bp in length
* 35291 35390: gap of 100 bp
* 35391 36199: contig of 809 bp in length
* 36200 36299: gap of 100 bp
* 36300 37156: contig of 857 bp in length
* 37157 37256: gap of 100 bp
* 37257 38097: contig of 841 bp in length
* 38098 38197: gap of 100 bp
* 38198 39015: contig of 818 bp in length
* 39016 39115: gap of 100 bp
* 39116 39895: contig of 780 bp in length
* 39896 39995: gap of 100 bp
* 39996 40786: contig of 791 bp in length
* 40787 40886: gap of 100 bp
* 40887 41686: contig of 800 bp in length
* 41687 41786: gap of 100 bp
* 41787 42549: contig of 763 bp in length
* 42550 42649: gap of 100 bp
* 42650 43459: contig of 810 bp in length
* 43460 43559: gap of 100 bp
* 43560 44293: contig of 734 bp in length
* 44294 44393: gap of 100 bp
* 44394 45225: contig of 832 bp in length
* 45226 45325: gap of 100 bp
* 45326 46138: contig of 813 bp in length
* 46139 46238: gap of 100 bp
* 46239 46987: contig of 749 bp in length
* 46988 47087: gap of 100 bp
* 47088 47853: contig of 766 bp in length
* 47854 47953: gap of 100 bp
* 47954 48758: contig of 805 bp in length
* 48759 48858: gap of 100 bp
* 48859 49664: contig of 806 bp in length
* 49665 49764: gap of 100 bp
* 49765 50569: contig of 805 bp in length
* 50570 50669: gap of 100 bp

```

Alignment Scores:

```

Pred. No.: 3 39e-45 Length: 146515
Score: 825.00 Matches: 161
Percent Similarity: 94.83% Conserves: 4
Best Local Similarity: 92.53% Mismatches: 7
Query Match: 23.78% Indels: 2
DB: 2 Gaps: 0

```

US-10-060-830-1114 (1-653) x AC025661 (1-146515)

```

QY 482 HisLeuSerProArgLysValThrThrValLeuGlnAlaAsp-SerAlaGluThrAlaGlu 501
DB 57222 AATCTCTTACTGTAATAATGACTTTCTGTTGTTGATATCCACAGATATGCTCA 57281
QY 501 nProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGlu 521
DB 57282 GCCACGCGTAGAGGAGATTTGTTGACATTCATCAAGATCTTAAACAGAGAGA 57341
QY 521 uGlyLysGluAlaGlyThrValAspLeuAspProTyrAspSerProGluGlnGluValTyr 541
DB 57342 AGGAAAGAGAGCAGCGATGACACCTAGATCTTACACTCAGCAGGAGGAGAGTTTA 57401
QY 541 rHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleLeu 561
DB 57402 TCATGCGCTATGCGAACAACATCCCAATTACGGGGCCGATATGCAACCCCAATCATCAT 57461
QY 561 tAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAla 581
DB 57462 GGCACATGTCAGGGCACCACCACTTCAGTTGTCAGCCCTCCACATCCACTTCAAGGC 57521
QY 581 aThrGlyAsnGlnProProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAs 601
DB 57522 TAGGGGACCAACCTCCCACTAGTGGAGACTTACAAATACACTTCTCCAGGACTGA 57581
QY 601 pSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGluLeuPr 621

```

```

DB 57582 CAGCTGCTCTCTAGCCAGCCAGCCAGTATGATVACCCGAAAGTGGAGCCAGGTCTAC 57641
QY 621 oAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlu 641
DB 57642 TGCCCCAGACGATTTGTTGTTACAGGTGCCACAGACACAGAGATTCAGAGAGAG 57701
QY 641 yArgAspGlyGluCys-AspValPheLysGluIleLeu 653
DB 57702 AAGGATGGGAGATGATGATGTTTAAAGAAATCCTT 57739

```

RESULT 14

```

AX118816 1464 bp DNA linear PAT 11-MAY-2001
LOCUS AX118816
DEFINITION Sequence 1 from Patent WO0129219.
ACCESSION AX118816
VERSION AX118816.1 GI:14035774
KEYWORDS
SOURCE

```

ORGANISM

```

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

REFERENCE

```

1 (bases 1 to 1464)
Donoho, G., Turner, C.A., Nehls, M., Friedrich, G., Zambrowicz, B. and
Sands, A.T.
Human cub-domain-containing protein and gene encoding the same
Patent: WO 0129219-A 1 26-APR-2001;
JOURNAL Lexicon Genetics Incorporated (US)

```

FEATURES

```

source
1..1464
/organism="Homo sapiens"
/db_xref="taxon:9606"
Location/Qualifiers

```

```

BASE COUNT 440 a 303 c 359 g 362 t
ORIGIN

```

Alignment Scores:

```

Pred. No.: 6.2e-43 Length: 1464
Score: 756.00 Matches: 181
Percent Similarity: 53.39% Conserves: 87
Best Local Similarity: 36.06% Mismatches: 158
Query Match: 21.79% Indels: 76
DB: 6 Gaps: 14

```

US-10-060-830-1114 (1-653) x AX118816 (1-1464)

```

QY 13 LeuLeuLeuGlnAspAlaGlyAlaGlnGlnGlyTyrCysGlyLeuGluMet 32
DB 139 CTCTCTTCCACCAAGCTCTTCAATGATATGTCCTACTGTGGA--AGTATGACTGTT 195
QY 33 AsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHis 52
DB 196 CCCAAGAACACTCTTGTGAACACAGAGTGAAGTAAACGTCGCTTGAAGATCCAC 255
QY 53 ValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThr 72
DB 256 ATTTCTGCGCGGGTTTGTGCTGACCTATGACGAGCAGCAGCCATCCAGATTAAATACA 315
QY 73 CysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGly 92
DB 316 TGTTTGGAGACGACTACCAATTATTGACAGACAGACAAATTTCTCCAGCTGGT 375
QY 93 CysLeuLeuProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerSer 112
DB 376 TGTAGACACTAGACAGACAGACATTTCTGGGAATATGTATGATATGATATGATACCTCT 435
QY 113 ProLeuGlyMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlnIle 132
DB 436 TTATTGTGCAAACTGCCATCCATGACGAGATATATGCTCATACACTAGTGGCCAGATC 495
QY 133 SerValValIleSerLysGlyIleProTyrTyrGlnSerSerLeuLeuAsnValThr 152
DB 496 AGTGTGCTCAGCGCAAGAGATCATGTCATATGAGGAGATTCTGGCCAAATGGTGTCTT 555

```

QY 153 SerValValGlyHisLeuSerThrSerLeuPheThrPheLeuThrSerGlyCysTyrGly 172
Db 556 TCGAGGGATGGTCCCTGTGACAGAGCATTTCTGTACCTCCAAATGGTTGACAGCAGA 615
QY 173 ThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeu 192
Db 616 TCCTTGAGTTTGAA-----CCTGACGGGCAAAATACAGAGCTTCTTCCCTCATGG 663
QY 193 GluTrpThrAspHisThrGlyClnGluAsnSerTrpLysProLysAlaArgLeuLys 212
Db 664 CAGTCGGTCAATCAGAGTGGACCAAGTTCACTGTCTCTCTGGCCCAAGCCGACTTCAG 723
QY 213 LysProGlyProProTrpAlaAlaPheAlaThrAspGlu-----TyrGlnTrp 228
Db 724 GACCAAGGCCCATCATGGCTTCGGCGACAGTACACCAACCAACACAGAGTGG 783
QY 229 LeuGlnIleAspLeuAsnLysLysLysLysLysLysLysLysLysLysLysLysLys 248
Db 784 CTGGAGATCGATTGGGGCGAGAAAGAAATAACAGGAATTAGGACCAACAGATCTACA 843
QY 249 MetValGluHisAsnTyrTrpValSerAlaTyrArgIleLeuTyrSerAspAspGlyGln 268
Db 844 CAGTCGAACCTCAACTTTATGTATAGAGTTTGTGATGAACCTCAAAACCAATATCT 903
QY 269 LysTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlnLys 288
Db 904 AAGTGAACACCTATNAAGCAATGTGAATATGAAGAAAGGTGTTTCAGGGTAACTCT 963
QY 289 AspTyrHisGlnAspValArgAsnAsnPheLeuProProIleAlaArgPheLeuArg 308
Db 964 AACTTTCCGGACCCAGTGCAGAAACAAATTTTCATCCCTCCATCGTGGCCAGATATGCGG 1023
QY 309 ValAsnProThrGlnTrpGlnLysLysLysLysLysLysLysLysLysLysLysLys 328
Db 1024 GTGTCCCGCCAGACATGGCCAGCAGAGATAGCCTTGAAGGTGGAGCTCATGTTGCCAG 1083
QY 329 PheIleProLysGlyArgProLysLeuThrGlnProProProProAsnSerAsn 348
Db 1084 -----ATTACACA-----GCTAAT 1098
QY 349 Asp-----LeuLysAsnThrThrAlaProProLysLysLysLysLysLysLys 364
Db 1099 GATTATTGTGTGGCGCAGACAGCAAGCAAGCAGCAGTGTTCACAACTAAGAAGAAGAT 1158
QY 365 ProLysPheThrGlnProLysGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 384
Db 1159 GAGACAATCACAAAGGCCATC-----CCCTCGGA 1188
QY 385 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 404
Db 1189 GAACA-----TCCACAGGATTAACATTAACAAG-----1218
QY 405 ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 424
Db 1219 GTGGCTATTCCATTGGTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1278
QY 425 IleLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysLysLysLysLysLysLys 444
Db 1279 GCCTTT-----AGAAAGAAGAAGAAGAAGT-----1308
QY 445 LeuProTyrTrp-----AspArgAlaGlyTyrTrpLysGlyMetLysGlnPhe 460
Db 1309 ---CCGTATGGATCAGCGGAGCTCAGAAACAGAGCTGTGGAGGAGATTAATAATATCCC 1365
QY 461 LeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSerSerGlu--- 479
Db 1366 TTT-----GCCAGACATCAGTCAGCTGAGTTT 1392
QY 480 ---ValAsnHisLeuSerProArgGluValThrValLeuGlnAlaAspSerAlaGlu 498
Db 1393 ACCATCAGCTATGATATGAGAGGAGATGACACAAAGTTAGATCTCATCAACAAGTAT 1452
QY 499 TyrAla 500

Db 1453 ATGGCA 1458
RESULT 15
AX118820
LOCUS AX118820 1620 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 5 from Patent WO0129219.
ACCESSION AX118820
VERSION AX118820.1 GI:14035776
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1620)
AUTHORS Donohue, G., Turner, C.A., Nehls, M., Friedrich, G., Zambrowicz, B. and Sands, A.T.
TITLE Human sub-domain-containing protein and gene encoding the same
JOURNAL Patent: WO 0129219-A 5 26-APR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
source 1..1620
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 456 a 353 c 422 g 389 t
ORIGIN
Alignment Scores:
Pred. No.: 7.03e-43 Length: 1620
Score: 756.00 Matches: 181
Percent Similarity: 53.39% Conservative: 87
Best Local Similarity: 36.06% Mismatches: 158
Query Match: 21.79% Indels: 76
Gaps: 14
DB:
US-10-060-830-1114 (1-653) x AX118820 (1-1620)
QY 13 LeuLeuLeuGluAspAlaGlyAlaGlnGlnGlyLysTyrCysGlyLeuGlyLeuGlnMet 32
Db 295 CTTCTCTTCCAGCAGCTCTTCAGATCAATATGCTCCATCTGTGA---AGTATGACTGTT 351
QY 33 AsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHis 52
Db 352 CCCAAGAAGACTCTTGTGTAACACAGTGAAGTAACCGCTTGTGAGAGTGTATGCCAC 411
QY 53 ValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThr 72
Db 412 ATTTCTGGCCGGGTTTTTTTGTGACCTATGCGAGCAGCAGCATCCAGATTTAATAACA 471
QY 73 CysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGly 92
Db 472 TGTGTAACAGAGCTAGCCATTTATTTGACAGACAGATACAGCAATTCGCCAGCTGGT 531
QY 93 CysLeuLeuProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSer 112
Db 532 TGTAGAGAGCTAGCAGGAGACATTTCTGGGAATATGATAGATGATATAGAGATACCTCT 591
QY 113 ProLeuCysMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIle 132
Db 592 TTATTGTGAAAGCTCCCATCCATGAGGAATAATTTGATGATGAAGTGTGAGTGTGAGATC 651
QY 133 SerValValIleSerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThr 152
Db 652 AGTGTCTCTCAGCGCAAGAGGATCACTCATATGAAGGGATTTCTGGCCATGTTCTT 711
QY 153 SerValValIleHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGly 172
Db 712 TCGAGGAGTGGTTCCTGTCTGACAGCAAGCGATTTCTGTTTACCTCCATGTTGACGAGA 771
QY 173 ThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeu 192
Db 772 TCCTTGAGTTTGA-----CCTGACGGGCAAAATCAGAGCTTCTTCTCATAGG 819

Qy	193	GlutThrThrAspHisThrGlyClnGlnuAsnSerTrpLysProLysLysAlaArgLeuLys	212
Db	820	CAGTCGGTCAATCAGAGGTGGACCAACAGTTCACTGGTCTCCTGGCCACCGCAGTTCAG	879
Qy	213	LysProGlyProProTirpAlaAlaPheAlaThrAspGlu-----TyrGlnTrp	228
Db	880	GACCAGGCCCATCATGGCTTCGGCGGCACAGTAGCAACACCAACACACAGAGTGG	939
Qy	229	LeuGlnIleAspLeuAsnLysGluLysLysLysLeThrGlyIleIleThrThrGlySerThr	248
Db	940	CTGGAGATCGATTTGGGGGAGAAAAAATAACAGGAATTAGGACACACAGGATCTACA	999
Qy	249	MetValClnHisAsnTyrTrpValSerIalaTyrArgIleLeuTyrSerAspAspGlyGln	268
Db	1000	CAGTCGAACCTCAACCTTTTATCTTAAGAGTTTGTGATGAACCTTCAAAAAACAATAATCT	1059
Qy	269	LysTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLys	288
Db	1060	AAGTGGAGACCTATAAAGGAATTGTGAATAATGAAGAAAAAGGTGTTTCAGGTGAACCTCT	1119
Qy	289	AspTyrHisGlnAspValArgAsnAsnPheLeuProProIleIleAlaArgPheIleArg	308
Db	1120	AACTTTGGGACCCAGTGTCAAAACAATTCATCCCTCCCACTGGCCACGATATGCGG	1179
Qy	309	ValAsnProThrClnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGln	328
Db	1180	GTTGTCCCCAGACATGTGCACGAGGATAGCTTTGAAGTGGAGCTCATTTGGTTGCCAG	1239
Qy	329	PheIleProLysGlyArgProProLysLeuThrGlnProProProArgAsnSerAsn	348
Db	1240	-----r-----ATTACACA-----GGTAAT	1254
Qy	349	Asp-----LeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla	364
Db	1255	GATTCAATTGTGTGGCGCAAGACAAGTCAAAACACAGTGTTCCTCAACTAAGAAAGAGAT	1314
Qy	365	ProLysPheThrClnProLeuGlnProArgSerSerAsnGluPheProAlaGlnThrGlu	384
Db	1315	GAGACAACTACAGGCCCATC-----CCCTCGGAA	1344
Qy	385	GlnThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp	404
Db	1345	GAAGA-----TCCACAGGAATAAACAATTACAAG-----1374	
Qy	405	ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu	424
Db	1375	GTGGCTATTCCATPTGTGTCTCTGTGTCTCTGGTGTGTGTTGCTGAATGGGATCTTTGCA	1434
Qy	425	IleLeuValCysAlaIaTrpHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrAsp	444
Db	1435	GCCTTT-----AGAAAGACAGACAGAAAGGAAGT-----1464	
Qy	445	LeuProTyrTrp-----AspArgAlaGlyTrpTrpLysGlyMetLysGlnPhe	460
Db	1465	---CCGTATGGATCAGCGGAGGCTCAGAAACACAGACTGTGTGAAGCAGATTAATAATCCC	1521
Qy	461	LeuProAlaLysAlaValAspHisGluGlnProValArgTyrSerSerSerGlu---	479
Db	1522	TTT-----GCCAGACATCATGCTACGTACGTGCTTT	1548
Qy	480	---ValAsnHisLeuSerProArgGluValThrValLeuGlnAlaAspSerAlaGlu	498
Db	1549	ACCATCAGCTATGTAATGAGAGGAGATGACACAAAAGTTAGATCTCATCAACAGTGAT	1608
Qy	499	TyrAla	500
Db	1609	ATGGCA	1614

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 21, 2003, 09:57:51 ; Search time 340.203 Seconds
(without alignments)
4322.586 Million cell updates/sec

Title: US-10-060-830-1114

Perfect score: 3470

Sequence: 1 MFLFLLLLVLILLLEDDAGA.....TQEVSGAGRGDCGVFKEIL 653

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10060830/runat_16012003_091558_29002/app_query_fasta_1.1742
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-LIST=45 -DOALIGN=200 -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=ptc -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-USER=US10060830 -CGCN_1_1_490 -ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-NO_XLPXY -NO_MMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3369	97.1	2046	21	AAZ51872 Human Factor 8 Hom
2	2340	67.4	2428	22	AAH34855 Human colon cancer
3	2126.5	61.3	1907	24	AAD35994 Human neuropilin-H
4	2126.5	61.3	1907	24	ABK4567 Human cDNA encoding
5	2126.5	61.3	1907	24	ABN59799 Novel human coding
6	1712.5	49.4	1388	24	ABK83640 Human cDNA differe
7	1039	29.9	594	23	AAZ80396 DNA encoding novel
8	897	25.9	640	21	AAZ80396 Human colon cancer
9	887.5	25.6	3594	22	AAZ80396 Human TANGO 229 CD
10	838	24.1	580	22	ABA60693 Human foetal liver
11	838	24.1	580	22	AAK08975 Human brain expres
12	838	24.1	580	22	AAK34866 Human bone marrow
13	838	24.1	580	22	AAI40582 Probe #9268 used t
14	828	23.9	467	22	ABA76819 Human foetal liver
15	828	23.9	467	22	AAK25439 Human brain expres
16	828	23.9	467	22	AAK51451 Human bone marrow
17	828	23.9	467	22	AAI57541 Probe #26227 used
18	756	21.8	1620	22	AAZ51872 Novel human protei
19	756	21.8	1761	22	AAZ51872 Novel human protei
20	756	21.8	1768	22	AAZ51872 Polynucleotide iso
21	416	12.0	1871	21	AAZ51872 Human neuropilin-H
22	378	10.9	1265	24	AAZ51872 Human cDNA encoding
23	378	10.9	1265	24	ABK45655 Murine factor V en
24	367	10.6	6585	21	AAZ51872 Human coagulation
25	357.5	10.3	6893	20	AAZ32182 Human brain cell s
26	355	10.2	2209	22	AAH57553 Human breast cance
27	355	10.2	4545	22	AAH26692 Human breast cance
28	355	10.2	4681	24	AAK86006 Human cDNA encoding
29	355	10.2	6909	21	AAZ51872 Human factor V CDN
30	353.5	10.2	6909	21	AAZ51872 Human Vth aggregat
31	353.5	10.2	6909	22	AAZ51872 Human DNA encoding
32	353.5	10.2	6909	22	AAZ51872 Gene #2275 used to
33	353.5	10.2	6925	17	AAZ51872 Human Factor V mut
34	353.5	10.2	6925	17	AAZ51872 Human Factor V CDN
35	353.5	10.2	6925	17	AAZ51872 Human del-1 encodi
36	352	10.1	1780	20	AAZ51872 Murine development
37	352	10.1	2303	18	AAZ51872 Murine del-1 encod
38	352	10.1	2303	20	AAZ51872 Human Factor-V wil
39	349.5	10.1	6909	16	AAZ51872 Gene #3021 used to
40	346	10.0	1270	24	ABN96523 HMFG 46 kDa antige
41	346	10.0	1384	16	AAO91199 Human developmenta
42	346	10.0	1780	18	AAZ51872 HMFG 46 kDa antige
43	346	10.0	1934	16	AAO91198 Human lactadherin
44	346	10.0	1934	21	AAZ51872 Human normal uteru
45	346	10.0	2015	20	AAZ51872

ALIGNMENTS

RESULT 1	AAZ51872
ID	AAZ51872 standard; cDNA; 2046 BP.
XX	AAZ51872;
AC	AAZ51872;
DT	04-JUL-2000 (first entry)
XX	Human Factor 8 Homologue cDNA.
DE	Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
XX	cerebroprotective; therapeutic; coagulation related disorder;
KW	haemophilia; stroke; screening; ss.
XX	Homo sapiens.
OS	
XX	
FH	Key
FT	CDS
	Location/Qualifiers
	19..2028

FT /*tag= a
 TT /product= "Factor 8 Homologue"
 XX

PN WO200012532-A1.

XX 09-MAR-2000.

XX 20-AUG-1999; 99WO-US19047.

XX 31-AUG-1998; 98US-0098521.

XX (ELIL) LILLY & CO ELI.

XX Rostock PRJ, Su W, Li XM;

XX WPI; 2000-256580/22.

DR P-PSDB; AAY70539.

XX Factor 8 homolog polypeptides and nucleic acids encoding them for
 PT treating coagulation related disorders such as hemophilia and stroke

XX Claim 1; Page 61-64; 68pp; English.

CC The present sequence is a cDNA encoding human Factor 8 homologue (F8H)
 CC which is a coagulation cofactor. The protein is selectively expressed in
 CC haematopoietic, heart and reproductive tissues. It has haemostatic and
 CC cerebroprotective activities. The F8H contains a Factor 5/8 signature
 CC and is useful as a therapeutic for treating coagulation related diseases
 CC such as haemophilia and stroke. The nucleic acid is useful as
 CC hybridisation probe and amplification primer for detecting deficiencies
 CC in the level of F8H mRNA, for screening F8H gene mutations and for
 CC monitoring regulation of gene expression. Fragments of the nucleic acid
 CC are also useful as diagnostic probes and primers, and can be used in
 CC screening methods such as those using DNA chips. The present sequence is
 CC also useful as a target to screen therapeutically useful modulators
 CC of the F8H.

SX Sequence 2046 BP; 615 A; 457 C; 467 G; 507 T; 0 other;

Alignment Scores:

Pred. No.: 8.7e-245 Length: 2046
 Score: 3369.00 Matches: 638
 Percent Similarity: 98.02% Conservative: 7
 Best Local Similarity: 96.96% Mismatches: 8
 Query Match: 97.09% Indels: 6
 DB: 21 Gaps: 1

US-10-060-830-1114 (1-653) x AA251872 (1-2046)

QY 1 MetProLeuPheLeuLeuLeuLeuValLeu-----LeuLeuLeuLeu 15
 DB 53 TTGCATTGAAGATTCTGATCTTGTCACCTTTAATTACTTGAGAATTTATATGGAATTG 112
 QY 16 GluAspAlaGlyAlaGlnGlnGlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSer 35
 DB 113 GAGTCAGCAGA-ACTGAATAGCAATACTGTGGTCTGGGGTTGCAATGAACCAATTCA 171
 QY 36 IleGluSerLysGlyAsnGluLeuThrLeuLeuPheMetSerGlyIleHisValSerGly 55
 DB 172 ATTTGAATCAAAAGGCAATGAATGAAATACATCTGTTTCATGAGTGAATCCATGTTCTGGA 231
 QY 56 ArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAsp 75
 DB 232 CGCGGATTTTGGCTCATACTCTGTATAGATAAACAAGATCTAATTACTTCTTTGGAC 291
 QY 76 ThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeu 95
 DB 292 ACTGCATCCCAATTTTGGAACTGAGTTCAGTAAAGTACTGCCAGCTGTTGCTGCTT 351
 QY 96 ProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerProLeuCys 115
 DB 352 CCTTTTCTGAGATATCTGGAACAATTTCTCATGGATATAGATTCTCGCCATTGTC 411

QY 116 MetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValVal 135
 DB 412 ATGGCTGGTGTGCATGCAGGAGTAGTCAACACAGCTGGCGGCCAAATCAGTGTGTGA 471
 QY 136 IleSerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThrSerValVal 155
 DB 472 ATTAGTAAAGGTATCCCTATTATGAAGTCTTTGGCTAACACGTCACATCTGTGGTG 531
 QY 156 GlyHisLeuSerThrSerLeuPheThrPheLeuSerGlyCysTyrGlyThrLeuGly 175
 DB 532 GGACACTTATCTACAAGTCTTTTACATTTAAGCAAGTGAATGTTATGGNACACTGGG 591
 QY 176 MetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluThrThr 195
 DB 592 ATGGAGCTCTGGTGTGCATCGCGGATCTCAAAATACACAGCATCTGCTGCTGAGTGGACT 651
 QY 196 AspHisThrGlyGlnGluAsnSerTyrLysProLysLysAlaArgLeuLysLysProGly 215
 DB 652 GACCACACAGGGCAAGAGAACAGTTGGAAACCCCAAAAAGCCAGGCTGNAAAAACCTGGA 711
 QY 216 ProProTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLys 235
 DB 712 CCGCCCTGGGCTGCTTTTGGCCTGATGATACAGTGGTGTACAAATAGATTGGAATAAG 771
 QY 236 GluLysLysIleThrGlyIleIleThrThrGlySerThrMetValGluHisAsnTyrTyr 255
 DB 772 GAAGAAGAAATACAGGCATTTAACCACCTGGATCCACCTGGTGGAGCACAAATCTAT 831
 QY 256 ValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTrpThrValTyrArgGlu 275
 DB 832 GTCTCTGCCTACAGAACTCTGACAGTGTGATGGCAGAAATGAGCTGTGTACAGAGAG 891
 QY 276 ProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspValArg 295
 DB 892 CCTGGTGTGGCAAGATAAGATATTCAAGGAAACCAAGATTATCACAGGATGTGGCT 951
 QY 296 AsnAsnPheLeuProProIleIleAlaArgPheIleArgValAsnProThrGlnTrpGln 315
 DB 952 AATAACTTTTGGCCACCAATTTATGCACGTTTATTAGAGTGAAATCTTACCCTGCGAG 1011
 QY 316 GlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgPro 335
 DB 1012 CAGAAAATTTGCCATGAAATGGAGCTGCTCGGATGTGAGTCTTATTCCTAAAGGTGCTCT 1071
 QY 336 ProLysLeuThrGlnProProProProArgAsnSerAsnAspLeuLysAsnThrThra 355
 DB 1072 CCAAAACCTTACTCAACCTCCACCTCCCTCGGAACCAAGATGACCTCAAAAACACTACGCC 1131
 QY 356 ProProLysIleAlaLysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSer 375
 DB 1132 CCTCCAAAATAGCCAAAGGTGCTGCCCAAAATTTACGCAACCACTACAACCTCGCAGT 1191
 QY 376 SerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThr 395
 DB 1192 AGCAATGAATTTCTGTCACAGACAGAAACAACACTGCCAGTCTCGATATCAGAAATAT 1251
 QY 396 ThrValThrProAsnValThrLysAspValAlaLeuAlaValLeuValProValLeu 415
 DB 1252 ACCGTAACCTCAAAATGTAAACCAAGATGTAGCGTGGCTGCGATCTTCTGCTCCCTGCTG 1311
 QY 416 ValMetValLeuThrThrLeuIleLeuValCysAlaThrPheHisTrpArgAsnArg 435
 DB 1312 GTCATGCTCTCACTACTCTCATTTCTCATATTAGTGTGCTGTGCTGCTGGCAAGAACAGA 1371
 QY 436 LysLysLysThrGluGlyThrTyrAspLeuProTyrTrpAspArgAlaGlyTrpTrpLys 455
 DB 1372 AAAAAAATACTGAAGGCACCTATGACTTACTTGGGACCGGCGAGGTGGTGGGAAA 1431
 QY 456 GlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGluThrProValArgTyr 475
 DB 1432 GGAATGAAGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAACCCCACTGCTGCTAT 1491
 QY 476 SerSerSerGluValAsnHisLeuSerProArgGluValThrThrValLeuGlnAlaAsp 495

Db	1492	ACACAGCAGGAAGTTAACTCCTGAGTCCAAAGAAAGTCACCCAGAGTGCCTGCAGGCTGAC	1551
Qy	496	SerAlaGluTyrAlaGlnProLeuValGlyGlyTleValGlyThrLeuHisGlnArgSer	515
Db	1552	TCGTCAGAGTAGCTCAGCCACTGGTAGGAGGAATTGTTGGTACACTTCATCAAGATCT	1611
Qy	516	ThrPheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSer	535
Db	1612	ACCTTTAAACCAAGAAAGGAAAGACAGGCTATGCAGACCTAGATATCCTTCAACAATCA	1671
Qy	536	ProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyr	555
Db	1672	CCAGGCGCAGGAAGTTATCATGCCCTATCTGAACACATCCCAATTCAGGGCGCTAGATAT	1731
Qy	556	AlaThrProIleMetAspMetSerGlyHisProThrThrSerValGlyGlnProSer	575
Db	1732	GCAACCCCAATCATCATGTGACATGTCAGGGCACCCCAACTTCAGTTGGTCAGCCCTCC	1791
Qy	576	ThrSerThrPheLysAlaThrGlyAsnGlnProProLeuValGlyThrTyrAsnThr	595
Db	1792	ACATCCACTTTCAAAGGCTACGGGGAAACCAAGCTCCTCCACCTAGTGGGAACCTTACAATACA	1851
Qy	596	LeuLeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAla	615
Db	1852	CTTCTCTCAGGAGCTGACAGCTGCTCCTCAGCCCGAGGCCCAAGTAGATACCCCGAAGACT	1911
Qy	616	GlyLysProGlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGln	635
Db	1912	GGGAGCCAGGCTACCTCCGCCAGACGCAATTTGGTGTACCAAGTGGCCACAGACGACACAA	1971
Qy	636	GluValSerGlyAlaGlyArgAspGlyGluCysAspValPheLysGluIleLeu	653
Db	1972	GAAGTATCAGGACGACGAAGGATGGGAATGCTGATGTTTTTAAAGAAATTCCTT	2025

RESULT 2	
AAH34855	
ID	AAH34855 standard; cDNA; 2428 BP.
XX	
AC	AAH34855;
XX	
DT	03-SEP-2001 (first entry)
XX	
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:1937.
XX	
KW:	Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW	colorectal carcinoma; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200122920-A2.
XX	
PD	05-APR-2001.
XX	
PF	28-SEP-2000; 2000WO-US26524.
XX	
PR	29-SEP-1999; 99US-0157137.
PR	03-NOV-1999; 99US-0163280.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;
XX	
WPI;	2001-235357/24.
DR	P-PSDB; AAG75450.
XX	
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptide
PT	useful for preventing, diagnosing and/or treating colorectal cancers
XX	
PS	Claim 1; Page 3444-3445; 9803pp; English.
XX	
CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where

The proteins are collectively known as colon cancer antigens. The col-
cancer antigens have cytostatic activity and can be used in gene
therapy and vaccine production. N and P may be used in the prevention,
diagnosis and treatment of diseases associated with inappropriate P
expression. For example, N and P may be used to treat disorders
associated with decreased expression by rectifying mutations or deletions
in a patient's genome that affect the activity of P by expressing
inactive proteins or to supplement the patients own production of P.
Additionally, N may be used to produce the colon cancer-associated Ps,
by inserting the nucleic acids into a host cell and culturing the cell
to express the proteins. N and P can be used in the prevention, diagnosis
and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
and AAH77789 represent sequences used in the exemplification of the
present invention.

N.B. Pages 666 to 682 and page 7053 of the sequence listing were
missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX

SQ Sequence 2428 BP; 677 A; 560 C; 535 G; 650 T; 6 other;

Alignment Scores:			
Pred. No.:	4,75e-167	Length:	2428
Score:	2340.00	Matches:	451
Percent Similarity:	84.9%	Conservative:	2
Best Local Similarity:	84.62%	Mismatches:	4
Query Match:	67.44%	Indels:	76
DB:	22	Gaps:	1

US-10-060-830-1114 (1-653) x AAH34855 (1-2428)

QY	1	MetProLeuPheLeuLeuLeuValLeuLeuLeuLeuGluAspAlaGlyValA	20
Db	75	ATGCCCTGTTCCTCCTGCTTACTTGTCCTCTCTGCTCGAGGACGGTGGAGCC	134
QY	21	GlnGln-----	22
Db	135	CAGCAAGGTCATGGATCTGGACACACTGTACTAGGCCTCGAGAGTGGAACCCTTACATCC	194
QY	22	-----	22
Db	195	ATAAACTACCACAGACCTATCCCAACAGCACACTGTTTTGTGAATGGGAGATCCGGTGAAG	254
QY	22	-----	22
Db	255	ATGGGAGAGAGATTCCCATCAAATTTGGTGACTTTTGACATTGAAGATTCTGATTTCTGT	314
QY	23	-----GlyLys 24	
Db	315	CACITTAATTACTTTGAGAATTTATATGGAATTTGGAGTCAGCAGCAACTGGAATAGGCAAA	374
QY	25	TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr	44
Db	375	TACTGTGGTCTGGGGTTGCAATGAACCATTTCAATGAATCAAAAGGCAATGAATATACA	434
QY	45	LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal	64
Db	435	TTGCTGTTCATGAGTGAATCCATGTTCTCGAGCCGGATTTTGGCCTCATACTCTGTT	494
QY	65	IleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu	84
Db	495	ATGATTAACACAGATCTAANTACTTGTGGACACTGCATCCAATTTTGTGAACCTCGAG	554
QY	85	PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIle	104
Db	555	TTCAGTAGGACTGCCCCAGCTGGTTGCTGCTTCCTTCCTTGTGAGATATCTGGACAAT	614
QY	105	ProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValVal	124
Db	615	CCTCATGGATATAGAGATTCTCCGCCATTTGTSCATGGCTGTGTCATGCGAGGAGTAGTG	674
QY	125	SerAsnThrLeuGlyGlyGlnIleSerValIleSerLysGlyIleProTyrTyrGlu	144
Db	675	TCAAACAGTTGGCGGCCAAATCAGTGTCTGTAATTAGTAAGGTATTCCCCATTATGAA	734

```
QY 145 SerSerLeuAlaAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThr 164
DB 735 AGTCTTTGGCTAACAGCTCACACTCTGTGGGACACTTATCTACAAGCTTTTACA 794
QY 165 PheLysThrSerGlyCysTrpGlyThrLeuGlyMetGluSerGlyValIleAlaSerPro 184
DB 795 TTTAAGACAAGTGGATGTTATGGACACTGGGATGGAGTGTGGTGTGATCGCGATCCT 854
QY 185 GlnIleThrAlaSerSerValLeuGluTrpThrAspHisThrGlyGlnGluAsnSerTrp 204
DB 855 CAAATAACAGCATCATCTGTGTGGAGTGGACTGACACAGGACAGGACAGAGTGG 914
QY 205 LysProLysLysAlaArgLeuLysLysProGlyProProTrpAlaAlaPheAlaThrAsp 224
DB 915 AAACCCAAAAGCCAGGCTGAAAAACCTGGACCCSCCTGGGCTGCTTTGCCACTGAT 974
QY 225 GluTyrGlnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThr 244
DB 975 GAATACCACTGGTGTAAATAGATTGTAATAGGAAAAAGAAATACAGGCAATTATAACC 1034
QY 245 ThrGlySerThrMetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSer 264
DB 1035 ACTGGAATCACCATTGGTGAGCACAAATTACTATGTGTCTGCTACAGAACTCTGTACAGT 1094
QY 265 AspAspGlyGlnLysTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePhe 284
DB 1095 GATGATGGCAGAAATGGACTGTGTACAGAGAGCTGTGTGGAGCAAGATAAGATATT 1154
QY 285 GlnGlyAsnLysAspTyrHisGlnAspValArgAsnAsnPheLeuProProlleIleAla 304
DB 1155 CAAGGAAACAAGATATATCACCAGGATGTGCGTAATATCTTTTGGCCACCAATTATTGCA 1214
QY 305 ArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeu 324
DB 1215 CGTTTATTAGTAGTGAATCTACCCAAATGGCAGCAAGAAATGGCCATGAAATGAGAGCTG 1274
QY 325 LeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProProPro 344
DB 1275 CTCGGATGTCACTTTATTCTTAAAGTGTCTCTCCCAAACTTACTCAACCTCCACTCCT 1334
QY 345 ArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla 364
DB 1335 CGGACAGCAATGACCTCAAAACACTACAGCCCTCCAAAAATAGCCAAAGGTGCTGCC 1394
QY 365 ProLysPheThrGlnProLysGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 384
DB 1395 CCATAATTACCAACCACTACAACTCGCAGTAGCAATGATTTCTTCGACAGACAGAA 1454
QY 385 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 404
DB 1455 CAAACAACCTGCCAGTCTGATATCAGAAATACTACCGTAACCTCAAAATGTAACCAAGAT 1514
QY 405 ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 424
DB 1515 GTAGCGCTGGCTGCAGTTCTTCTCCCTGTGTGGTGTGTCATGCTCTCACTCTCATCTTC 1574
QY 425 IleLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrAsp 444
DB 1575 ATATTAGTGTGTGTGGCTGGCACTGGGAAACAGAAAGAAAGAAAGAAAGTAAAGGCACTATGAC 1634
QY 445 LeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMet 457
DB 1635 TTACCTTACTGGACCGGCAAGTAACTCACTCACTGCTGCTT 1673
RESULT 3
ID AAD35994
XX AAD35994 standard; DNA; 1907 BP.
AC AAD35994;
XX
XX 09-AUG-2002 (first entry)
XX
```

Human neuropilin-Hy2 DNA.

Human; neuropilin-like polypeptide; neuropilin-Hy2; neovascularisation; neurodegenerative disease; Alzheimer's disease; learning; angiogenesis; thrombocytopaenia; memory; platelet; plastic anaemia; antiinflammatory; ophthalmological; paroxysmal nocturnal haemoglobinuria; protozoacide; wound healing; tissue repair; Parkinson's disease; Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome; trauma; spinal cord; cerebrovascular; stroke; ulcer; immune; infection; multiple sclerosis; human immunodeficiency virus; HIV; autoimmune disorder; dermatological; systemic lupus erythematosus; rheumatoid arthritis; antithyroid; cancer; autoimmune thyroiditis; diabetes mellitus; graft-versus-host disease; myasthenia gravis; autoimmune inflammatory eye disease; gene therapy; neurotropic; neuroprotective; vulnery; anticonvulsant; antiparasitic; cerebroprotective; tranquilliser; virucide; antibacterial; cytostatic; immunosuppressive; chromosome 6q21; gene; ds.

Homo sapiens.

Key	Location/Qualifiers
CDS	101..1258
	/*tag= a
	/product= "Human neuropilin-Hy2 protein"
	/note= "This region designated as SEQ.ID.NO:8 is specifically referred in claim 1"
sig_peptide	101..160
	/*tag= b
mat_peptide	161..1255
	/*tag= c
	/product= "Human mature neuropilin-Hy2 protein"

WO200222815-A1.

21-MAR-2002..

12-SEP-2001; 2001WO-US28488.

11-SEP-2000; 2000US-0659671.

06-SEP-2001; 2001US-317902P.

(HYSE-) HYSEQ INC.

Tang YT;

WPI; 2002-393966/42.

P-PSDB; AAE22716.

Novel isolated human Neuropilin-Hy1 and Neuropilin-Hy2 polypeptides useful for treating neurodegenerative diseases e.g. Alzheimer's disease, and for diagnosing and mapping genetic neuronal defects -

Claim 1; Page 126-128; 152pp; English.

The invention relates to neuropilin-like polypeptide (neuropilin-Hy1 and neuropilin-Hy2) and their corresponding nucleic acids. The neuropilin-like polypeptides and polynucleotides are useful in modulating neuronal growth regenerative capacity, treating neurodegenerative diseases, diagnosing and mapping genetic neuronal defects and degenerative diseases like Alzheimer's disease and for treating learning and memory disorders. They are also useful for inducing angiogenesis, neovascularisation, as well as organ growth and development e.g. heart and other tissues. Antagonists of neuropilin-like polypeptides are useful for treating cancer and other malignant diseases. Neuropilin is used to treat platelet disorders e.g. thrombocytopaenia, plastic anaemia and paroxysmal nocturnal haemoglobinuria and is used in nerve tissue growth or regeneration, in wound healing, tissue repair and replacement and in healing of bones, incisions and ulcers. Compositions comprising the sequences of the invention are useful for treating diseases of peripheral nervous system e.g. Parkinson's disease, Huntington's disease, Shy-Drager syndrome, amyotrophic lateral sclerosis, and traumatic disorders e.g. spinal cord disorders, head trauma and cerebrovascular diseases e.g. stroke, ulcers, immune deficiencies and immune disorders, infections by hepatitis viruses, human immunodeficiency virus (HIV), herpes viruses,

Db	881	TCAAAATAACAGCATCACT- 	GTGCTGGATGGGACTGACCACACAGCGGCAGAGAACAGTTG	939
Qy	204	pLys-ProLysLysAlaAarqLeuLys 	LysProGlyProPtoPtoAlaAlaAlaAlaThra	224
Db	940	GAAACCCAAAAAGCCAGGCTGA 	AAAACCTGGACCGCCTTGGT-GCTTTGGCCACTG	998
Qy	224	spGluTyrrGlntPrleuGlnIleasp 	LeuAnsLysGluLysLysllethrGlylleller	244
Db	999	ATCAATTAACAGTGTTTACAATTA 	GATATTGTAATTAAGCAAAAAGAAATTAACAGGCAATTATAA	1058

Pred. No.:	4.5e-151	Length:	1907
Score:	2136.50	Matches:	446
Percent Similarity:	83.33%	Conservative:	4
Best Local Similarity:	82.59%	Mismatches:	7
Query Match:	61.28%	Indels:	88
DB:	24	Gaps:	21

Qy	1 MetProLeuPheLeuLeuLeuValLeuLeuLeuGluAspAlaGlyAla 20
Db	101 ATGCCTCGTTCPCCTGGCTCTTAAGTGCTGTCTGTCTGAGACGCTGGAGCC 160

Db 161 CAGCAAGGTGATGGATGTGGACACACTGTACTAGGCCCTGAGAGTGGAACCTTACATCC 220

Qy 22 22

Db · 341 CACTTTAATTACTTGAGAA TTATATGGAA TTGGAGTCAGCAGAACTGAAATAGGCAA 400

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Qy 45 LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal 64

Db 521 ATAGATAACAAGATCTAATTACTTGTTTGGACACTGCATCCAAATTTTGGAACTGAG 580

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

QV 105 ProHisGlyTyrArgAspSerProLeuCysMetalaglvvalHisAlaglvval 124

[illegible]

Db 701 TCAACACGTTGGGGCGCCAAATCAGTGTGTAAATTAGTAAAGGTATCCCCATTATGAA 760

QY I43 SEI SEI LEUATASASINVAIT IUI SEI VA I VA I GY I H I S B E N S E I I I I I I S E I L E U P I N E I I O I

165 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757

DD TTTAAGACAAGTGGATGTTATGGAAACATGGGGATGGAGTCTGGTGTGATCGCGGATCC 880

Sequence 1907 BP; 548 A; 449 C; 406 G; 504 T; 0 other;

QY	244	hrThrGlySerThrMetValGluHisAsnTyrValSerAlaTyrArgIleLeuTyrS	264
DB	1059	CCACTGGATCCACCATGGT - GAGCACAAATTACTATGTCTGCCTACAGAAATCCTGTACA	1117
QY	264	erAspAspGlyGlnLysTrpThrValTyrArgGluProGly - ValGluGlnAspLysIle	283
DB	1118	GTGATGATGGGCAGAAATGGAGCTGTGTACAGAGAGCCCTGGCTGGAGCAAGATAAGATA	1177
QY	284	PheGlnGlyAsn - LysAspTyrHisGlnAsp - ValArgAsnAsnPheLeuProProIle	302
DB	1178	TTTTCAGGAAACAAAGAATATATCACCAGGATGGTGGCTAATACTTTTGGCCAACTT	1237
QY	303	IleAlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMet	322
DB	1238	ATTGCACG - TTTATTAGAGTGAATCCTACCAATGGCAGCAGAAATTTGCATGAAATG	1296
QY	323	GluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGlnProPro	342
DB	1297	GAGCTGCTCGGATGTCACTTATTCCTAAAGGTGCTCTCAAACCTACTCAACCTCCA	1356
QY	343	ProProArgAsnSerAsnAspLeuLysAsnThrAlaProProLys - - - IleAlaLys	361
DB	1357	CCTCTCTCGAACAGCAATGACCTCAA - AACACTACAGCCCTCCAAAAATTTAGCCAAA	1415
QY	362	GlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAla	381
DB	1416	GGTGTGTGCCCAAAATTTACGCAACCACTACAACCTCGCAGTAGCAATGATTCCTGCA	1475
QY	382	GlnThrGluGlnThrAlaSerProAspIleArgAsnThrThrValThrProAsnVal	401
DB	1476	CAGCAGACAACAACTGCCAGTCCCTGATATCAGAAATACTCCGTAACCTCCAAATGTA	1535
QY	402	ThrLysAspValAlaLeuAlaValLeuValProValLeuValMetValLeuThrThr	421
DB	1536	ACCAAGATGTAGCGCTGCCTCACTTCTTGTCCCTGTGTGGTCATGGCTCCTCACTACT	1595
QY	422	LeuIleLeuIleValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGluGly	441
DB	1596	CTCATTTCTATATTAGTGTGCTTTGGCACTGGAGAACACAGAAAGAAAAAACTGAAGGC	1655
QY	442	ThrTyrAspLeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMet	457
DB	1656	ACCTATGACTTACCTTACTGGGACCGGCAGGTACTCACTGGTCTT	1703
RESULT	5		
ID	ABN59799		
CDNA:	1907 BP		

1
2
3
4
5
6

RESULT 5	
ABN59799	
ID	ABN59799 standard; cDNA; 1907 BP.
XX	
AC	ABN59799;
XX	
DT	28-JUN-2002 (first entry)
XX	
DE	Novel human coding sequence SEQ ID NO: 210.
XX	
KW	Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
KW	antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW	neuroprotective; antiparkinsonian; protein therapy; Est;
KW	expressed sequence tag; gene; ss.
XX	
OS	Homo sapiens.
PN	WO200222660-A2.
XX	
PD	21-MAR-2002.
XX	
Pf	10-SEP-2001; 2001WO-US26015.
XX	
PR	11-SEP-2000; 2000US-0659671.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao Q&, Ren F;

PI	Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX	
XX	WPI; 2002-292408/33.
DR	P-PSDB; ABB97386.
XX	
XX	An isolated polynucleotide for treating diseases associated with its
PT	encoded polypeptide such as cancer and multiple sclerosis -
PT	
XX	Claim 1; SEQ ID NO 210; 509pp; English.
PS	
XX	
CC	The present invention provides the protein and coding sequences of 444
CC	novel human proteins. These were isolated from expressed sequences tags
CC	(ESTs). They can be used to stimulate cell growth, to regulate
CC	haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC	e.g. in burn treatment, to regulate the immune system e.g. to treat
CC	multiple sclerosis, to regulate the immune system e.g. to treat
CC	multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC	infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC	stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC	e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC	Parkinson's disease. The present sequence is a coding sequence of the
CC	invention.

Sequence 1907 BP: 548 A: 449 C: 406 G: 504 T: 0 other: XX
S0

Alignment Scores:

Pred. No.:	4.5e-151	Length:	1907
Score:	2126.50	Matches:	446
Percent Similarity:	83.33%	Conservative:	4
Best Local Similarity:	82.59%	Mismatches:	7
Query Match:	61.28%	Indels:	88
DB:	21	Gaps:	2
US-10-060-830-1114 (1-653) x ABNS9799 (1-1907)			

US-10-060-830-1114 (1-653) x ABN59799 (1-1907)

Qy	1	MetProLeuPheLeuLeuLeuLeuValLeuLeuLeuLeuGluAspAlaGlyAla	20
Db	101	ATGCCTCTGTTCCTCCGCTCTTACTTGCTCCTCTCTCGAGACGCTGGAGCC	160
Qy	21	GlnGln-----	22
Db	161	CAGCAAGGTGATGGATGTGAACACACTGTACTAGGCCCTGAGAGTGGAAACCCTTACATCC	220
Qy	22	-----	22
Db	221	ATAAACAACCACAGACACCTATCCCAACAGCACACTGTTTGTGAATGGGAGATCCGTGTAAG	280
Qy	22	-----	22
Db	281	ATGGGAGAGAGAGTTGCATCAAATTTGGTGACTTTTGACATTGAAGATTCTGATCTTGT	340
Qy	23	-----GlyLys 24	
Db	341	CACHTTAATTACTTGAGAATTTATAATGGAATTTGGAGTCAGCAGAACTGNAATAGGCANA	400
Qy	25	TyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGluIleThr	44
Db	401	TACTGTGCTGGSGTTGCAATGAACCATTTCAATTGAATCAAAGAGCAATGAAATCACA	460
Qy	45	LeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerVal	64
Db	461	TTGCTGTTCATGAGTGGAAATCCATGTTTCTGGACGGCGGATTTTTGGCCCTCATACTCTGTT	520
Qy	65	IleAspLysGlnaspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGlu	84
Db	521	ATAGATAACANGATCTAAATCTGTTTGGACACTGCATCCCAATTTTTTGGAACTTGAG	580
Qy	85	PheSerLysTyrCysProAlaGlyCysLeuLeuProPheAlaGluIleSerGlyThrIle	104
Db	581	TTCACTAAGTACTGCCACGCTGGTTGTCTCTCTCTTCTGAGATATCTCGAACAAAT	640
Qy	105	ProHisGlyTyrArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValVal	124
Db	641	CCTCATGGATATAGAGATTCTCGCCAATTGTCATGCTGTGTCATGCAGAGAGTAGT	700

CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1388 BP; 432 A; 289 C; 302 G; 364 T; 1 other;

Alignment Scores:

Pred. No.: 5,21e-120 Length: 1388
 Score: 1712.50 Matches: 374
 Percent Similarity: 91.45% Conservative: 11
 Best Local Similarity: 88.84% Mismatches: 19
 Query Match: 49.35% Indels: 23
 DB: 24 Gaps: 4

US-10-060-830-1114 (1-653) x ABR83640 (1-1388)

QY 1 MetProLeuPheLeuLeuLeuLeuValLeu-----LeuLeuLeuLeu 15
 Db 148 TTGCATTTGAAGATTCTGATTTGTCACCTTTAATTACTTGAGAAATTTAATGAAGATTG 207
 QY 16 GluAlaGlyAlaGlnGlnGlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSer 35
 Db 208 GAGTCAGCAGA-ACTGAATAGCAAAATCTGTGGTCTGGGGTGGCAATCAACCATTC 266
 QY 36 IleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGly 55
 Db 267 ATTGAATCAAAGCAATGAATCACATTCCTCTCATGAGTGAATCCATGTTCTGGA 326
 QY 56 ArgGlyPheLeuAlaSerTyrSerValIleAspLysGln-AspLeuIleThrCysLeuAs 75
 Db 327 CGCGGATTTTGGCCCTCATCTCTGTATAGATAAACAAGATCTAATTACTGTTTGA 386
 QY 75 pThrAlaSerAsn-pheLeuGluProGluPheSerLysTyrCys-ProAlaGlyCysLeu 94
 Db 387 CACTGCATCCAAATTTTGGACCTGAGTTCAGTAGTACTGCCCGAGCTGGTGTCTG 446
 QY 95 LeuProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerProLeu 114
 Db 447 CTTCCTTTCTCAGATATCTGAACAATTCCTCATGGATATAGAGATTCTCTGCCATTG 506
 QY 115 CysMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerVal 134
 Db 507 TGCATGGCTGGTGTGATGAGAGTAGTGTCAACACGTTGGCGGCCCAATCAGTGT 566
 QY 135 ValIleSerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThrSerVal 154
 Db 567 GTAATTAGTAAGGTATTCCTATTATGAAGTTCCTTGGCTAACACGTCACATCTGTG 626
 QY 155 ValGlyHisLeuSerThrSerLeu---PheThrPheLysThrSerGlyCysTyrGlyThr 173
 Db 627 GTGGGACACTTGATACACAGTCTTTTACATTTAAGACAAGTGGATGTATGAACA 686
 QY 174 LeuGlyMetGluSerGly-ValIleAlaAspProGlnIleThrAlaSerValLeuG1 193
 Db 687 CTGGGNATGGAGTCTGGTGTGATCGCGGATCTCAAAATACAGCATCCT-GTGCCTGGA 745
 QY 193 uTrpThrAspHisThrGlyGlnGluAsnSerTrpLys-ProLysLysAlaArgLeuLysL 213
 Db 746 GTGGACTGACACACAGCGGCAAGACACTTGAACCCCAAAAAGCCAGGCTGAAA 805
 QY 213 ysProGlyProProTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspL 233
 Db 806 AACCTGGACCGCTGGGT-GCTTTTCCCACTGATGAATACCACTGGTGTACAAATAGATT 864
 QY 233 euAsnLysGluLysLysIleThrGlyIleIleThrThrGlySerThrMetValGluHisA 253
 Db 865 TGAATAGSAAAGAAATAACAGGCATTTAACCCTGGATCCACCATGCT-GAGCACA 923
 QY 253 snTyrTrpValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTrpThrValT 273
 Db 924 ATTACTATGTCTGCTCAGAAATCTGTACAGTATGATGGCAGAAATGGACTGTGT 983
 QY 273 yrArgGluProGly-ValGluGlnAspLysIlePheGlnGlyAsn--LysAspTyrHisG 292
 Db 984 ACAGAGACCTGGTCTGGGAGCAGAGATAAGATATTTCAAGGAACAAGAAATATACCC 1043

QY 292 lnAsp-valArgAsnAsnPheLeuProIleIleAlaArgPheIleArgValAsnPro 311
 Db 1044 AGGATGGTGGTAATAACTTTTGGCCACCAATATTTCACAG-TTTATTAGAGTAATCCT 1102
 QY 312 ThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheIlePro 331
 Db 1103 ACCCAATGGCAGCAGAAATATTCATGAAATGAGCTGCTCGATGTCAATTTATTCCT 1162
 QY 332 LysGlyArgProProLysLeuThrGlnProProProProProProArgAsnSerAsnAspLeuLys 351
 Db 1163 AAAGGTGCTCTCCAAAATTTACTCAACCTCCACCTCTCGGAACAGCAATGACCTCAA- 1221
 QY 352 AsnThrThrAlaProProLys---IleAlaLysGlyArgAlaProLysPheThrGlnPro 370
 Db 1222 AACACATACAGCCCTCCAAAAATTTAGCCAAAGGTGCTGCCCAAAATTTACGCAACCA 1281
 QY 371 LeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerPro 390
 Db 1282 CTACAACCTCGCAGTAGCAATGAATTCCTGTCACAGACACACAAACAACACTGCCAGTCT 1341
 QY 391 AspIleArgAsnThrThrVal---ThrProAsnValThrLysasp 404
 Db 1342 GATATCAGAAATTAACCTACCGTAAACTCCAAATGTAACCAAGAT 1386

RESULT 7

AAS65485

ID AAS65485 standard; cDNA; 594 BP.

XX AAS65485;

AC AAS65485;

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #1289.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

OS WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

XX P-PSDB; ABG01298.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 1; SEQ ID No 1289; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC Imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 594 BP; 175 A; 160 C; 139 G; 120 T; 0 other;

Alignment Scores:
 Pred. No.: 1.17e-69 Length: 594
 Score: 1039.00 Matches: 197
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 29.94% Indels: 0
 DB: 23 Gaps: 0

US-10-060-830-1114 (1-653) x AAS65485 (1-594)

Qy 457 MetLysGlnPheLeuProAlaLysAlaValAspHisGluThrProValArgTyrSer 476
 Db 1 ATGAAGCAGTTCTCTCGTGAAGAGCAGTGGACCATGAGGAAACCCAGTTCGTATAGC 60
 Qy 477 SerSerGluValAsnHisLeuSerProArgGluValThrValLeuGlnAlaAspSer 496
 Db 61 AGCAGCGAAGTTATACCTACCTGAGTCCAGAGAGAGTCCACAGTGCCTGACGCTCT 120
 Qy 497 AlaGluThrAlaGlnProLeuValGlyGlyLeuValGlyThrLeuHisGlnArgSerThr 516
 Db 121 GCAGAGTATGCTCAGCCACCTGGTAGGAGGAATTGTTGGTACACTTCATCAAGATCTACC 180
 Qy 517 PheLysProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerPro 536
 Db 181 TTTAAACCAAGAGAGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
 Qy 537 GlyGlnGluValTyrHisAlaTyrAlaGluProLeuProLeuThrGlyProGluTyrAla 556
 Db 241 GGGCAGGAAGTTTATCATGCTATGCTGCAACCACTCCCAATTCAGGGGCGCTGAGTATGCA 300
 Qy 557 ThrProIleLeuMetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThr 576
 Db 301 ACCCAATCATCATGAGATGTCAGGCGACCCACCAACTTCAGTGTGTCAGCCCTCCACA 360
 Qy 577 SerThrPheLysAlaThrGlyAsnGlnProProProLeuValGlyThrTyrAsnThrLeu 596
 Db 361 TCCACTTTCAGGCTACGGGGAACCAACCTCCCCACTAGTGGGAACCTTACAATACACTT 420
 Qy 597 LeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGly 616
 Db 421 CTCCTCCAGGACTGACAGTGTCTCTCAGCCAGGCGCCAGTATGATACCCGGAAGCTGGG 480
 Qy 617 LysProGlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGlu 636
 Db 481 AAGCCAGGTCATCTGCTGCCAGACCAATTTGTTGTTACCAAGGTGCCACAGACGACACAA 540
 Qy 637 ValSerGlyAlaGlyArgAspGlyGluCysAspValPheLysGluIleLeu 653
 Db 541 GTATCAGGAGCAGGAGGATGGGCAATGTGATGCTTTTAAAGAAATCCCTT 591

RESULT 8

AZ80396/c

ID AZ80396 standard; cDNA; 640 BP.

XX AC AZ80396;

XX XX 07-APR-2000 (first entry)

XX DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:480.

XX Human; gene expression product; diagnosis; tumour; colon cancer;
 KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
 KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
 KW hyperplasia; ds.
 XX Homo sapiens.
 XX WO9964576-A2.
 XX PD 16-DEC-1999.
 XX 09-JUN-1999; 99WO-IB01062.
 XX PF 10-JUN-1998; 98US-0088801.
 XX (FARB) BAYER CORP.
 XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
 PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
 PI Schlegel R;
 XX WPI; 2000-087220/07.
 XX Novel nucleic acids, used to develop products for the diagnosis and
 PT treatment of disorders involving unwanted cell proliferation,
 PT particularly cancers, especially colon cancer .
 XX Claim 15; Page 322; 469pp; English.
 XX AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
 CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
 CC cDNA clones can be used to generate antisense oligonucleotides which
 CC can be used for antisense therapy. Methods and products from the present
 CC invention can be used for identifying and/or classifying cancerous cells
 CC present in a human tumour, particularly in solid tumours, e.g.
 CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
 CC can be used for developing agents for the diagnosis and treatment of
 CC disorders involving unwanted cell proliferation, such as neoplasia,
 CC dysplasia or hyperplasia.
 XX SQ Sequence 640 BP; 148 A; 134 C; 160 G; 192 T; 6 other;

Alignment Scores:

Pred. No.: 6.67e-59 Length: 640
 Score: 897.00 Matches: 194
 Percent Similarity: 91.98% Conservative: 1
 Best Local Similarity: 91.51% Mismatches: 17
 Query Match: 25.85% Indels: 6
 DB: 21 Gaps: 0

US-10-060-830-1114 (1-653) x AAZ80396 (1-640)

Qy 299 LeuProProIleLeuAlaArgPheIleArgValAsnProThrGlnTrpGlnLysIle 318
 Db 630 TTGCCACCAATTATGGCAGC-TTTATTAGAGT-AATCCTTACCCTATGCCAGCAGAAATT- 574
 Qy 319 AlaMetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProLysLeu 338
 Db 573 GCCATGAAAT-GAACTGCTCGGATGTCAGTTTATTCNTAAAGT-NGTCNTCAAAACTT 516
 Qy 339 ThrGlnProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProLys 358
 Db 515 ACTCAACTCCACCTCTTNGNACAGCAATGNCCTCAAAACCACTACAGCCCTCCAAA 456
 Qy 359 IleAlaLysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGlu 378
 Db 455 TA-GCCAAAGGTCTGTCGCCCAAAATTTACGCAACCACTACAACTTCGCAATGAA 397
 Qy 379 PheProAlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThr 398
 Db 396 TTTCTCTGCAGACAGAAACAACTGCCAGTCTCTATATCATGAAATATCTACCGTAAC 337

QY 399 ProAsnValThrLysAspValAlaLeuAlaValLeuValProValLeuValMetVal 418
 DB 336 CCAATGTAACCAAGATGAGCGCTGGCTGCAGTCTTGTCCCTGGTGGTCATGGTC 277
 QY 419 LeuThrThrLeuLeuLeuLeuValCysAlaThrPheHisTrpArgAsnArgLysLysLys 438
 DB 276 CTCACACTCTCATCTTCATATTAGTGTGTGGTGGCACTGGAGAAACAGAAAGAAAAA 217
 QY 439 ThrGluGlyThrTrpAspLeuProTyrTrpAspArgAlaGlyTrpTyrLysGlyMetLys 458
 DB 216 ACTGAAGGCACCTATGACTTACCTTACTGGGGCGGCGGAGGTGGTGGAAAGGATGAAG 157
 QY 459 GluPheLeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSerSerSer 478
 DB 156 CAGTTTCTCTCTGCAAAAGCAGTGCCACCATCAGGAGAAACCCAGTTCGTATAGCAGCAG 97
 QY 479 GluValAsnHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGlu 498
 DB 96 GAAGTTAATCACCCTGATGCTCAAGAGAAAGTACCACAGTGTGTCAGGCTGACTGTCAGAG 37
 QY 499 TyrAlaGlnProLeuValGlyGlyValGlyThr 510
 DB 36 TATGCTCAGCCACTGCTAGGAGGAATGTTGGTACC 1

RESULT 9

AA000660
 ID AA000660 standard; cDNA; 3594 BP.

AC AA000660;

XX 07-SEP-2001 (first entry)

XX Human TANGO 229 cDNA.

XX Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
 KW skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
 KW bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
 KW anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
 KW malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
 KW attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
 KW hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
 KW muscular dystrophy; immuno-competence; vertebrate; blood; serum.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 72..2219

FT /*tag= a

FT /product= "Human TANGO 229"

FT sig_peptide 72..173

FT /*tag= b

FT mat_peptide 174..2216

FT /*tag= c

FT /product= "Mature human TANGO 229"

XX WO200129088-A1.

XX 26-APR-2001.

XX 23-JUN-2000; 2000WO-US17386.

XX 19-OCT-1999; 99US-0420707.

XX (MILL-) MILLENNium PHARM INC.

XX Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;

XX WPI; 2001-308477/32.

XX P-PSDB; AAU00670.

XX New isolated nucleic acid molecule for diagnosis, prevention, and
 PT therapy of human and other animal disorder, or as modulating agent for
 PT regulating cellular processes

XX

PS Claim 1; Fig 1; 263pp; English.

XX The sequence represents a cDNA which encodes human TANGO 229 polypeptide.
 CC This protein and similar others exhibit the ability to affect growth,
 CC proliferation, survival, differentiation, activity, morphology, or
 CC movement/migration of, e.g. T cells and cells of the heart, liver,
 CC pancreas, placenta, brain, lung, skeletal muscle, kidney, spleen, lymph
 CC node, peripheral blood leukocyte, bone marrow or thymus tissue. They can
 CC be used as modulating agents for regulating cellular processes, thus, the
 CC proteins and their associated nucleic acids can be used to prognosticate,
 CC prevent, diagnose, or treat disorders associated with physiological
 CC processes. These disorders include abnormal blood coagulation, asthma,
 CC anaphylaxis, hepatitis, multiple sclerosis, cancer, coronary artery
 CC disease, malaria, atopic dermatitis, amyotrophic lateral sclerosis,
 CC meningitis, attention deficit disorder, Crohn's disease, gastroenteritis,
 CC goitre, hypoglycaemia, diabetes mellitus, endometriosis, pulmonary
 CC embolism and muscular dystrophy. Antibodies to disorders such as these
 CC can be made by providing a polypeptide of the invention to an
 CC immuno-competent vertebrate and harvesting blood or serum from the
 CC vertebrate.

XX Sequence 3594 BP; 958 A; 810 C; 921 G; 905 T; 0 other;

Alignment Scores:

Pred. No.: 3,09e-57 Length: 3594
 Score: 887.50 Matches: 230
 Percent Similarity: 50.76% Conservative: 103
 Best Local Similarity: 35.06% Mismatches: 221
 Query Match: 25.58% Indels: 102
 DB: 22 Gaps: 21

US-10-060-830-1114 (1-653) x AA000660 (1-3594)

QY 13 LeuLeuLeuGluAspAlaGlyAlaGlnGlnGlyLysTyrCysGlyLeuGlyLeuGlnMet 32
 DB 366 CTCTCTTCCACCACTCTTCAGATCAATATGCTCCATCTACTGTGA---AGTATGACTGT 422
 QY 33 AsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHis 52
 DB 423 CCCAAGAAGCTCTTTGTAACACAAAGTCAAGTCAACCGTCCGCTTTGAGAGTGGATCCAC 482
 QY 53 ValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThr 72
 DB 483 ATTTCTGGCGGGGTTTTTTTGTGCTGACCATGCGAGCAGCACCACATCCCAATTTAATAACA 542
 QY 73 CysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCysProAlaGly 92
 DB 543 TGTGGAAACGAGCTAGCCATTATTTGAAGACAGATACAGCAAAATTCGCCAGCTGGT 602
 QY 93 CysLeuLeuProPheAlaGluIleSerGlyThrIleProHisGlyTyrArgAspSerSer 112
 DB 603 TGTAGACGCTAGCAGGAGACATTTCTGGGAATATGTTAGATGGATATAGATACCTCT 662
 QY 113 ProLeuCysMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIle 132
 DB 663 TTATTGTCAAAAGCTGCCATCCATCGAGGAATAATTTGCTAGTAAGTAGTGGCCAGATC 722
 QY 133 SerValValIleSerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThr 152
 DB 723 AGTGTGCTTTCAGCGCAAGGATCATGTCGATATGAAGGATTTCTGGCAATGGTGTCTT 782
 QY 153 SerValValGlyHisLeuSerThrSerLeuPheThrPhePheSerGlyCysTyrGly 172
 DB 783 TCGAGGGATGTTCTCCCTGTACAGCAAGCGATTTCTGTGTACCTCCAAATGGTTCAGCAGA 842
 QY 173 ThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeu 192
 DB 843 TCCTTGAGTTTGA-----CCTGACGGGCAAAATCAGAGCTTCTTCTCATGG 890
 QY 193 GluTrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaGlyLeuLys 212
 DB 891 CAGTCGGTCAATGAGAGTGGAGACCAAGTCTACTGTCTCTCTGCGCCAGCCGACTTCAG 950

```
Qy 213 LysProGlyProProTrrpAlaAlaPheAlaThrAspGlu-----TyrGlnTrp 228
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 951 GACCAAGGCCCATCATGGCTTCGGCCACAGTAGCAACACCAACACAGAGTGG 1010
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 229 LeuGlnIleAspLeuAsnLysGluLysLysThrGlyIleIleThrThrGlySerThr 248
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1011 CTGGAGATCGATTGGGGGAGAGAAAGAAATAACAGAAATTAGGACACAGAGATCTACA 1070
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 249 MetValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspAspGlyGln 268
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1071 CAGTCGAACCTCAACTTTATGTTAAGAGTTTGTGAGTAAGTTTCAAAAACAATAATCT 1130
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 269 LysTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLys 288
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1131 AAGTGGAGACCTATTAAGGAATTGTGAATAATCAAGAAAGGTGTTTCAGGGTAACCTCT 1190
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 289 AspTyrHisGlnAspValArgAsnAsnPheLeuProProlIleIleAlaArgPheIleArg 308
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1191 AACTTTCGGGACCCAGTCGAACAAATTCATCCCTCCCATCGTGGCCAGATATGTGGG 1250
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 309 ValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGln 328
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1251 GTTGTCCTCCACAGATGACACAGGATAGCCTTGAAGTGGAGCTCATTTGGTGGCCAG 1310
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 329 PheIleProLysGlyArgProProLysLeuThrGlnProProProArgAsnSerAsn 348
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1311 -----ATTACAAA-----GGTAAT 1325
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 349 Asp-----LeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla 364
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1326 GATTCAATTGGTGTGGCGCAAGCAAGTCAAGACACCAAGTGTTCACAAAGAAAGAGAT 1385
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 365 ProLysPheThrGlnProLysGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 384
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1386 GAGACAATCACAGGCCCATC-----CCCTCGAA 1415
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 385 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 404
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1416 GAAACA-----TCCACAGGAATAAATCAACG----- 1445
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 405 ValAlaLeuAlaAlaValLeuValProValLeuValMetValLeuThrThrLeuIleLeu 424
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1446 GTGGCTATTCCATGCTGCTCTGTTGCTGCTGTTGCTGGAATGGGATCTTTGCA 1505
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 425 IleLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGluGlyThrTyrAsp 444
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1506 GCCTTT-----AGAAAGAAAGAAAGAAAGAGT----- 1535
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 445 LeuProTyrTrp-----AspArgAlaGlyTrpTrpLysGlyMetLysGlnPhe 460
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1536 ---CGTATGGATCAGCAGAGGCTCAGAAACAGACTTTGGAAGCAGATTAATATATCC 1592
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 461 LeuProAlaLysAlaValAspHisGluGluThrProValArgTyrSerSerGlu--- 479
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1593 TTT-----GCCAGACATCATCTAGCTAGCTGAGTTT 1619
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 480 ---ValAsnHisLeuSerProArgGluValThr-----ThrValLeuGlnAlaAsp 495
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1620 ACCATCAGCTATGTAANTAGAGAGAGATGACACAAAAGTTAGATCTCATCACAGTGAT 1679
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 496 SerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSer 515
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1680 ATGGCAGATTAACACAGCCCTCATGATTGGCACCGGACAGATCAGCAGGAGGCTCC 1739
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 516 ThrPheLysPro-----GluGluGlyLysGluAlaGlyTyrAlaAspLeuAspPro--- 532
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1740 ACCTTCGGGCCCATGGACACGATCCGAGGAGGAGG---GTGAGCAGCCGATGCCGCC 1796
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 533 -----TyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIle 550
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1797 GGCCACTATGACTGCCCGCAGCGCGCGGCCGCCAGAGTAGCGCTGCCCTCGCGCCC 1856
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Qy 551 ThrGlyProGluTyrAlaThrProIleIleMetAspMetSerGlyHisProThrThrSer 570
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1857 CCGAGCCCGGACGTACGCCACGCCCATCGTG-----GAGCGCAGCTGCTGGCGGCC 1907
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 571 ValGlyGlnProSerThrThrThrPheLysAlaThrGlyAsnGlnPro----- 586
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1908 CACACGTTCTCTGCGCAGAGCGGCTACCGCTCCCGAGGCGCCAGCCGCCCAACAC 1967
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 587 -----ProProLeu-----ValGlyThrTyrAsnThrLeuLeu 597
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1968 TCCCTCTCCCGCGCGGCTTCTCCCGTAGCGGTGGGGCCCGCAGGAGACTAT 2027
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 598 SerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAla-Glyly 617
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2028 CAAAGGCCACACGCGCACGCTCGGACAGAGGCTACGACCGCGCCCAAGCTGTCTCAGC 2087
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 617 sProGlyLeuProAlaProAspGluLeuValTyrGlnValProGln 632
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2088 GCCTCTCCCGCAGGAGCGGACCTCTCAGAAAGCCCCCA 2133
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 10
ABA60693
ID ABA60693 standard; DNA; 580 BP.
XX
AC ABA60693;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #8998.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN W0200157277-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 28-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 1; SEQ ID NO 8998; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;
XX
Alignment Scores: 1.67e-54 Length: 580
Pred. No.:
```

Score: 838.00 Matches: 161
 Percent Similarity: 95.38% Conservative: 4
 Best Local Similarity: 93.06% Mismatches: 7
 Query Match: 24.15% Indels: 1
 DB: 22 Gaps: 0

US-10-060-830-1114 (1-653) x ABA60693 (1-580)

Qy 482 HisLeuSerProArgGluValThrValLeuGlnAlaAsp-SerAlaGluTyrAlaGl 501
 Db 21 AATCTCTTTACTAGTAAGTAACTTTCTGTTGTTGTCATATCCACAGAGTATGCTCA 80
 Qy 501 nProLeuValGlyGlyLeuValGlyThrLeuHisGlnArgSerThrPheLysProGluGl 521
 Db 81 GCCACTGGTAGGAGGAATGTTGGTACACTTCATCAAGATCTACCTTTAAACCAGAAGA 140
 Qy 521 uGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTy 541
 Db 141 AGGAAAGAAGCAGGCTATGCACACTAGATCCTTACAACTCACCAGGGCAGGAAGTTTA 200
 Qy 541 rHisAlaTyrAlaGluProLeuProLeuProLeuThrGlyProGluTyrAlaThrProLeuLeMe 561
 Db 201 TCATGCTATGCTGAACCACTCCCAATACGGGGCCTGAGTATGCAACCCCAATCATCAT 260
 Qy 561 tAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAl 581
 Db 261 GGACATGTTCAGGGCACCACCAACTTCAGTTGGTGCAGCCCTCCACATCCACTTTCAAGGC 320
 Qy 581 aThrGlyAsnGlnProProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAs 601
 Db 321 TAGGGGAACCAACTCCCCCACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGA 380
 Qy 601 pSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuPr 621
 Db 381 CAGCTCTCTCAGCCAGCCAGTATGATACCCGAAAGCTGGGAAGCCAGGCTCTACC 440
 Qy 621 oAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGl 641
 Db 441 TGCCCCAGAGCAATTTGGTGTACAGGTGCCACAGAGCACACAAGAAGTATCAGGAGCAGG 500
 Db 501 AAGGGATGGGAATGTGATGTTTTTAAAGAAATCCTT 537

RESULT 11

AAK08975

ID AAK08975 standard; DNA; 580 BP.

XX

AC AAK08975;

XX

DT 05-NOV-2001 (first entry)

XX

DE Human brain expressed single exon probe SEQ ID NO: 8966.

XX

KW Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KW epilepsy; cancer; ss.

XX

OS Homo sapiens.

XX

PN W0200157275-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US000667.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-483446/52.

XX

PT Single exon nucleic acid probes for analyzing gene expression in human brains -

XX

PS Example 4; SEQ ID NO: 8966; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.

XX

SQ Sequence 580 BP; 169 A; 150 C; 125 G; 136 T; 0 other;

Alignment Scores:

Pred. No.: 1.67e-54 Length: 580
 Score: 838.00 Matches: 161
 Percent Similarity: 95.38% Conservative: 4
 Best Local Similarity: 93.06% Mismatches: 7
 Query Match: 24.15% Indels: 1
 DB: 22 Gaps: 0

US-10-060-830-1114 (1-653) x AAK08975 (1-580)

Qy 482 HisLeuSerProArgGluValThrValLeuGlnAlaAsp-SerAlaGluTyrAlaGl 501
 Db 21 AATCTCTTTACTAGTAAGTAACTTTCTGTTGTTGTCATATCCACAGAGTATGCTCA 80

Qy 501 nProLeuValGlyGlyLeuValGlyThrLeuHisGlnArgSerThrPheLysProGluGl 521
 Db 81 GCCACTGGTAGGAGGAATTTGGTACACTTCATCAAGATCTACCTTTAAACCAGAAGA 140

Qy 521 uGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTy 541
 Db 141 AGGAAAGAAGCAGGCTATGCACACTAGATCCTTACAACTCACCAGGGCAGGAAGTTTA 200

Qy 541 rHisAlaTyrAlaGluProLeuProLeuProLeuThrGlyProGluTyrAlaThrProLeuLeMe 561
 Db 201 TCATGCTATGCTGAACCACTCCCAATACGGGGCCTGAGTATGCAACCCCAATCATCAT 260

Qy 561 tAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAl 581
 Db 261 GGACATGTTCAGGGCACCACCAACTTCAGTTGGTGCAGCCCTCCACATCCACTTTCAAGGC 320

Qy 581 aThrGlyAsnGlnProProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAs 601
 Db 321 TACGGGAACCAACTCCCCCACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGA 380

Qy 601 pSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuPr 621
 Db 381 CAGCTCTCTCAGCCAGCCAGTATGATACCCGAAAGCTGGGAAGCCAGGCTCTACC 440

Qy 621 oAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGl 641
 Db 441 TGCCCCAGAGCAATTTGGTGTACAGGTGCCACAGAGCACACAAGAAGTATCAGGAGCAGG 500

Qy 641 YArgAspGlyGluCysAspValPheLysGluLeuLeu 653
 Db 501 AAGGGATGGGAATGTGATGTTTTTAAAGAAATCCTT 537

RESULT 12

AAK34866

ID AAK34866 standard; DNA; 580 BP.

XX

AC AAK34866;

Qy	482	HisLeuSerProArgGluValThrValLeuGlnAlaasp-SerAlaGluTyrAlaG1	501
Db		: : : : : : : : :	
Qy	21	AATCTCTTACTAGTAAATGACTTTCTGCTTTTGTGATATCCACAGAGTATCCTCA	80
Qy	501	nProLeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheIysProGluG1	521
Db	81	GCCACTGGTAGGAGGAATTGTTGGTACACTTCATCAAGATCTACCTTTAAACCAAGA	140
Qy	521	uGlyLysGluAlaGlyTyrAlaaspLeuaspProFyrAsnSerProGlyGlnGluValTy	541
Db	141	AGGAAGAAGCAGGGCTATGCAGACCTAGATCCTTACAACCTCACCGGCGAGGAAGTTTA	200
Qy	541	rHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMe	561
Db	201	TCATGCCTATGCTGAACCACTCCCAATACGGGGGCGCTGAGTATGCAACCCCAATCATCAT	260
Qy	561	tAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAl	581
Db	261	GGACATGTCAGGGCACCCACAACTTCAGTTGGTCAGCCCTCCACATCCACTTTCAGGCG	320

US-10-060-830-1114 (1-653) x AAI40582 (1-580)

QY 482 HisLeuSerProArgGluValThrThrValLeuGlnAlaAsp-SerAlaGluTyrAlaG1 501
Db 21 AATCTCTTTTACGTAGTAAATGACTTCTTGGTTGTTGATATATACACAGAGTATGCTCA 80
QY 501 nProLeuValGlyGlyValGlyThrLeuHisGlnArgSerThrPhelYsProGluG1 521
Db 81 GCCACTGGTAGAGAGNATTTGGTACACTTTCATCAAGATCTACCTTTAAACAGAGAGA 140
QY 521 uGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTy 541
Db 141 AGGAAAAGAGCAGGCTATGCAGACCTAGATCCCTTCAACTCACCAGGCGAGGAAGTTTA 200
QY 541 rHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleIleMe 561
Db 201 TCATGCTATGCTGAACCACTCCCAATTACGGGGCCCTGAGTATGCAACCCCAATCATCAT 260
QY 561 tAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPhelYsAl 581
Db 261 GGACATGTCAGGGCACCACCACTTCAGTTGGTCAGCCCTCCACATCCACTTTCAGGCG 320
QY 581 aThrGlyAsnGlnProProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAs 601
Db 321 TACGGGGAACCAACCTCCCCCACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGA 380
QY 601 pSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuPr 621
Db 381 CAGCTCTCTCTCAGCCAGCCAGATGATATACCCCGAAAGCTGGGAAGCCAGGTCTACCC 440
QY 621 oAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaG1 641
Db 441 TGCCCCAGACGAAATGGTGTACAGGTGCCACAGAGACACACAGAGTATCAGGAGGAGG 500
QY 641 yArgAspGlyGluCysAspValPhelYsGluIleLeu 653
Db 501 AAGGGATGGGAATGTGATGTTTTTAAAGAAATCCTT 537

RESULT 14
ID ABA76819 standard; DNA; 467 BP.
XX
AC ABA76819;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #25124.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human fetal liver -
Claim 4; SEQ ID NO 25124; 639pp + sequence listing; English.
The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human foetal liver. The present sequence is a single exon nucleic acid probe of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 467 BP; 137 A; 128 C; 107 G; 95 T; 0 other;

Alignment Scores: 7.19e-54 Length: 467
Pred. No.: 828.00 Matches: 155
Score: 828.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.86% Indels: 0
DB: 22 Gaps: 0
US-10-060-830-1114 (1-653) x ABA76819 (1-467)

QY 499 TyrAlaGlnProLeuValGlyGlyValGlyThrLeuHisGlnArgSerThrPhelYs 518
Db 3 TATGCTCAGGCACCTGGTAGGAGGAAATTTGGTACACTTTCATCAAGATCTACCTTTAA 62
QY 519 ProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGln 538
Db 63 CCAGAAGAGGAAAGAGCAGGCTATGCAGACCTAGATCCTTACAACCTCACCAGGCGAG 122
QY 539 GluValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrPro 558
Db 123 GAAGTTTATCATGCTATGCTGTGAACCACTCCCAATTACGGGGCCCTGAGTATGCAACCCCA 182
QY 559 IleIleMetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThr 578
Db 183 ATCATATGGACATGTGAGGGCACCACCACTTCAAGTTGGTGGCCCTCCACATCACT 242
QY 579 PhelYsAlaThrGlyAsnGlnProProProLeuValGlyThrTyrAsnThrLeuLeuSer 598
Db 243 TTCAAGGCTAGCGGGAACCAACCTCCCCCACTAGTGGGAACCTTACAATACACTTCTCTCC 302
QY 599 ArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysPro 618
Db 303 AGGACTGACAGCTGCTCTCAGCCAGGCCAGTATGATACCCGAAAGCTGGGAAGCCCA 362
QY 619 GlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSer 638
Db 363 GGTCTACCTGCCCCAGAGCAATTTGGTACAGGTGCCACAGAGCACACAGAAAGTATCA 422
QY 639 GlyAlaGlyArgAspGlyGluCysAspValPhelYsGluIleLeu 653
Db 423 GGACAGAGGAGTGGGATGTGATGTTTTTAAAGAAATCCTT 467

RESULT 15
AAK25439
ID AAK25439 standard; DNA; 467 BP.
XX
AC AAK25439;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 25430.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.

QY 639 GlyAlaGlyArgAspGlyGluCysAspValPheLysGluIleLeu 653
 | | | | | | | | | | | | | | | | | |
Db 423 GGACAGNAGGGATGGGAATGTGATGTTTTTAAAGAAATCCTT 467

US-10-060-830-1114 (1-653) x AAK25439 (1-467)

Qy	499	Tyr	Ala	Gln	Pro	Leu	Val	Gly	Gly	Ile	Val	Gly	Thr	Leu	His	Gln	Arg	Ser	Thr	Phe	Lys	518	
Db	3	TAT	GCT	CAG	CCACT	TGG	TAG	GAG	GAAT	TGT	TGG	TAC	ACT	TCAT	CAA	GAT	ACT	TAC	CTTT	ATAA	62		
Qy	519	Pro	Glu	Glu	Gly	Lys	Glu	Ala	Gly	Tyr	Ala	Asp	Leu	Asp	Pro	Tyr	Asn	Ser	Pro	Gly	Gln	538	
Db	63	CC	AGA	AGA	GAA	AGA	GAG	CAG	GCT	TAT	G	CAG	ACC	TAG	AT	TCC	TAC	AACT	CAC	AGG	CAG	122	
Qy	539	Glu	Val	Tyr	His	Ala	Tyr	Ala	Glu	Pro	Leu	Pro	Ile	Thr	Gly	Pro	Glu	Tyr	Ala	Thr	Pro	558	
Db	123	GA	AGTTT	TAT	CAT	GCT	TAT	GCT	GA	ACC	CAC	TCC	CAAT	TAC	GGG	CCCT	GAG	TAT	GCA	ACC	CCA	182	
Qy	559	Ile	Ile	Met	Asp	Met	Ser	Gly	Ile	His	Pro	Thr	Thr	Ser	Val	Gly	Gln	Pro	Ser	Thr	Ser	Thr	578
Db	183	AT	CAT	CAT	GAC	ATG	T	CAG	GGC	ACC	CC	CAACT	TCA	GTT	TGG	T	GCG	CCCT	C	CAC	ATC	CCACT	242
Qy	579	Phe	Lys	Ala	Thr	Gly	Asn	Gln	Pro	Pro	Pro	Leu	Val	Gly	Thr	Tyr	Asn	Thr	Leu	Leu	Ser	598	
Db	243	TT	CAAGGCT	ACG	GGGA	CAAC	CAACT	CCCC	CACT	AGT	TGG	GAAC	TAC	CAAT	AC	ACT	TCT	CTCC	302				
Qy	599	Arg	Thr	Asp	Ser	Cys	Ser	Ser	Ala	Gln	Ala	Gln	Tyr	Asp	Thr	Pro	Lys	Ala	Gly	Lys	Pro	618	
Db	303	AG	GACT	GAC	AGCT	GTCT	CCT	CAG	CCCG	AGG	CCCA	TAT	GAT	AT	CCCC	AGAA	AGCT	GGGA	AGCCA	362			
Qy	619	Gly	Leu	Pro	Ala	Pro	Asp	Glu	Leu	Val	Tyr	Gln	Val	Pro	Gln	Ser	Thr	Gln	Glu	Val	Ser	638	
Db	363	GG	TCT	ACCT	TGCC	CCAG	CAANT	TGGT	GTAT	CC	AGSGT	GCC	CAG	CAG	CAC	ACA	AGA	GAT	ATCA	422			

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 23, 2003, 12:48:26 ; Search time 2321.88 Seconds
(without alignments).
4554.775 Million cell updates/sec

Title: US-10-060-830-1114
Perfect score: 3470
Sequence: 1 MFLFLLLLVLLVLLLELDAGA.....TOEVSAGRGDCGVFKEIL 653

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-O=/cgn2_1/USPTO.spool/US10060830/runat_23012003_124821_6284/app_query.fasta_1.1742
-MODEL=frame+p2n.model -DEV=xlh
-DB=EST_QPWT-fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=DIOSUM62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THRT=SCORE-ppt -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pt0 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830.ecgn_1.1.3526_eunat_23012003_124821_6284 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_estl: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_man: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1494.5	43.1	1065	13	BM450083 AGENCOURT
2	1236	35.6	694	13	BI860608 603386804
3	1212	34.9	712	9	AI378788 tc67b04.x
4	1151.5	33.2	1272	11	AK006805 Mus muscu
5	1090.5	31.4	641	9	AI435602 th79c10.x
6	1087	31.3	680	14	BQ014739 UI-H-EDI-
7	1064	30.7	618	12	BE894226 601437167
8	1047.5	30.2	671	9	AA147037 zc32a05.s
9	1023	29.5	643	10	BB179946 BB179946
10	1011	29.1	664	10	AW383889 QV3-HT036
11	990	28.5	600	14	BM781107 MLN_4_E0
12	977	28.2	608	10	BB657747 BB657747
13	974	28.1	575	10	AV603144 AV603144
14	974	28.1	575	12	BF167024 601774167
15	972	28.0	628	14	BQ037529 BQ037529
16	956	27.6	560	12	BF074573 BF074573
17	940	27.1	535	9	AI417256 tg76f11.x
18	898	25.9	575	10	AW383902 QV3-HT036
19	880	25.4	587	10	AW383890 QV3-HT036
20	877	25.3	538	14	N21309 YX53q04.sl
21	861.5	24.8	1061	14	BM905706 AGENCOURT
22	791	22.8	457	9	AI276772 q164h03.x
23	780	22.5	525	9	AI290500 q198d11.x
24	778.5	22.4	689	9	AJ444823 AJ444823
25	778.5	22.4	764	9	AJ441638 AJ441638
26	778	22.4	459	14	N46066 YV35q05.r1
27	772.5	22.3	809	9	AJ445175 AJ445175
28	770	22.2	463	12	BF777080 288892 MA
29	756	21.8	574	13	BQ32248 BQ32248
30	753	21.7	778	9	AJ441380 AJ441380
31	750.5	21.6	817	9	AJ396784 AJ396784
32	747	21.5	439	14	H99543 YX29q09.r1
33	729	21.0	455	13	BI132538 AR032C05L
34	725	20.9	466	9	AI750960 cno6g08.x
35	725	20.9	584	9	AL674742 AL674742
36	710	20.5	422	9	AL046164 DFE2P434B
37	689	19.9	395	9	AI242761 q118h05.x
38	645.5	18.6	462	10	BB854159 BB854159
39	607	17.5	532	14	BQ775634 UI-H-PH0-
40	606	17.5	693	12	BF732384 naella12.
41	604.5	17.4	908	12	BG178371 602330141
42	593.5	17.1	459	12	BG748212 602705836
43	589	17.0	416	13	BI119166 AR024H09P
44	589	17.0	597	9	AJ450411 AJ450411
45	559	16.1	437	14	H80004 yS65d12.r1

ALIGNMENTS

RESULT 1
BM450083

LOCUS
DEFINITION

BM450083 1065 bp mRNA linear EST 05-FEB-2002
AGENCOURT_6393434 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5528102
5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1065)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12204 row: g column: 15
High quality sequence stop: 665.
Location/Qualifiers
1..1065
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5528102"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2 kb. Library constructed by Life Technologies."

BASE COUNT 312 a 280 c 236 g 237 t

ORIGIN

Alignment Scores:
Pred. No.: 2,75e-138 Length: 1065
Score: 1494.50 Matches: 284
Percent Similarity: 98.28% Conservative: 1
Best Local Similarity: 97.93% Mismatches: 4
Query Match: 43.07% Indels: 1
DB: 13 Gaps: 1

US-10-060-830-1114 (1-653) x BM450083 (1-1065)

QY 363 ArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGluPheProAlaGln 382
DB 1 CQTGCCCAAAATTTACGCAACCATACAACTCCAGTAGCAATGAAATTCCTGCACAG 60
QY 383 ThrGluGlnThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThr 402
DB 61 ACAGACACAAACACCTGCCAGTCTCTATACAGATACTACCGTAACTCCAAATGTAC 120
QY 403 LysAspValAlaLeuAlaValLeuValProValLeuValMetValLeuThrLeu 422
DB 121 AAAGATGTAGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 423 IleLeuIleLeuValCysAlaTrpHisTrpArgAsnArgLysLysLysThrGluGlyThr 442
DB 181 ATTCTCATATTAGTGTGCTGGCACTGGAGAAACAGAAAGAAAGAAAGAAAGAAAG 240
QY 443 TyrAspLeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuPro 462
DB 241 TATGACTTACCTTACCTGCGCCGGGAGGTGGTGGAAAGGAATGAAGCAGTTCTCTCT 300
QY 463 AlaLysAlaValAspHisGluThrProValArgTyrSerSerSerGluValAsnHis 482
DB 301 GCAAAAGCAGTGGACCATGAGGAACCCCGCTGCTATAGCAGCAGCAGCAAGTTATAC 360
QY 483 LeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGluThrAlaGlnPro 502
DB 361 CTGAGTCCAAAGAGAAGTCCACAGTGGCTGCGAGGCTGCTGCGAGTATGCTCAGCCA 420
QY 503 LeuValGlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGly 522
DB 421 CTGCTAGGAGGAATTTGGTACACTTCATCAAGATCTACCTTTAAACAGAGAGGA 480
QY 523 LysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGlnGluValTyrHis 542
DB 481 AAAGAAGCAGGCTATGACAGACCTAGATCCTTACAACTCACCAGGCGAGGAAGTTATCAT 540

QY 543 AlaTyrAlaGluProLeuProLeuProfileThrGlyProGluTyrAlaThrProfileMetAsp 562
DB 541 GCCTATGCTGAACCACTCCCAATTTACGGGGCTGAGTATGCAACCCCAATCATCATGCAC 600
QY 563 MetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThr 582
DB 601 ATGTGAGGGCACCACCAACTTCAGTTGGTGGTCCCTCCACATCCATCCATTCACGGCTACG 660
QY 583 GlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAspSer 602
DB 661 GGGAAACCACTCCCTCCACTAGTGGGAACCTTACATACATCTCTCTCCAGGACTGACAGC 720
QY 603 CysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuProAla 622
DB 721 TGCTCTCCAGCCAGCCAGTATGATACCCGGAAGCTGGGAAGCAGGCTTACCTGCC 780
QY 623 ProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArg 642
DB 781 CCAGACGAATTTGGTGTACCAAGTCCCGCAGAGCACCAGCAAGATATCCGGAGCAGGAAG 840
QY 643 Asp---GlyGluCysAspValPheLysGlu 651
DB 841 GGATCGGGGAATGTGAATGTTTAAAGAA 870

RESULT 2

B1860608
LOCUS B1860608 694 bp mRNA linear EST 10-OCT-2001
DEFINITION B03386804F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395609 5', mRNA sequence.

ACCESSION B1860608

VERSION B1860608.1 GI:16001355

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 694)

NTN-MGC http://mgc.nci.nih.gov/.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

Contact: Robert Strausberg, Ph.D.

COMMENT

Tissue Procurement: DCTD/DRP
Email: cgabs-r@mail.nih.gov
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

High quality sequence stop: 692.

FEATURES

source
Location/Qualifiers
1..694

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5395609"

/clone_lib="NIH_MGC_87"

/tissue_type="mammary adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally; oligo-dr primed.

Average insert size 1.383 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Library."

BASE COUNT 230 a 165 c 147 g 152 t

ORIGIN

Alignment Scores:
Pred. No.: 8,89e-113 Length: 694
Score: 1236.00 Matches: 228
Percent Similarity: 99.57% Conservative: 2
Best Local Similarity: 98.70% Mismatches: 1
Query Match: 35.62% Indels: 0

Db 244 ----- 243
QY 253 NYVSAVRLYSDDGQKWTYVREGEVDKIFQGNKYHODVRNLFPIIARFIRVNPT 312
Db 244 ----- 246
QY 313 QWQKIAMKMLGCGQFIPKGRPKLTPQPPRNSDLKNTAPPKIAKGRAPKFTQPLQ 372
Db 247 ----- 246
QY 373 PRSSNEFFAQTEQTASPDIRNTTVPNTKDVLAALVPLVMVLTLLILVCANHW 432
Db 247 ----- 272
QY 433 RNRKKTEGTYDLPY-----WDRAGWKGKQFLPAKAVDHEETP-VRYSSSE--VNHLSR 485
Db 273 --RKRKKGN--PVSDAQKTCGCKQIKY-----PEARHSTFTSYDNE 315
QY 486 REVT---TVLQADSLEYAQLVGGIVGLTHQBSTKP-----EE-----GKEAGYADLPY 533
Db 316 KEMTKLDTSDMADYQOPLMIGTGVARKGSTPRMDTDEEVRVNTEASGH-----Y 370
QY 534 NSPGQEVHVAEPLIPITGPEVATPII-----NDMSGHPTTSVGPSTSTFKATG 583
Db 371 DCPHPRGRHEVALPHTHSEPEYATPIVERHLLRAHTFTSQSGIRVPG-PRPTHEHSHSG 429
QY 584 NQPPPLVGTNTLLSRDSCSSAQAYDTPKAGKPLPAPDELVYQVQSTQSVGAGRD 643
Db 430 GPPATGATOVESYQRPASPKVGGYDKP-AASSFLDSRD-----PASQSOIMTSGDD 482
QY 644 G 644
Db 483 G 483

RESULT 4

US-10-115-563-14
; Sequence 14, Application US/10115563
; Publication No. US20030008307A1
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H
; TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
; C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
; AND COMPOSITIONS THEREOF
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; Patent Counsel
; STREET: 10666 No. US20030008307A1th Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/115,563
; FILING DATE: 02-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,488
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: 449.0
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312

; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2224 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-115-563-14

Query Match 10.0%; Score 347; DB 9; Length 2224;
Best Local Similarity 45.8%; Pred. No. 1.8e-19;
Matches 76; Conservative 21; Mismatches 59; Indels 10; Gaps 3;
QY 166 KTSYGCYGLMESGVADPQITASSVLE--WTDHTGQENSWKPKARLKKPG--PPWAAF 221
Db 2062 EYNGSTPLGMEKIKENKQITASSFKKSWGDI-----WEPFARLNAQGRVNAWQAK 2115
QY 222 ATDEYQWLIDLNKKKTKITITGSTVMVHNYVSAYRILYSDDGQKWTYVREGEVD 281
Db 2116 ANNKQWLEIDLLKKTKITAITQCKSLSSSEMYKSYTIHYSEQGVENKPYRLKSSMYD 2175
QY 282 KIFGKNDYHODVRNLFPIIARFIRVNPTQWQKIAMKMLLGC 327
Db 2176 KIFGNTNTKGHVKNFENPPIISRFIRVPKTNQISITLRLEFGC 2221

RESULT 5

US-10-003-132-6
; Sequence 6, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROPTILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-003-132-6

Query Match 9.3%; Score 324; DB 9; Length 458;
Best Local Similarity 20.1%; Pred. No. 1.3e-18;
Matches 140; Conservative 77; Mismatches 159; Indels 322; Gaps 22;
QY 2 PLFLLLLVLLLELDAGAGQKVC-----LGLQNHST-----ESK 39
Db 8 PSVLALLFVCAPLRLOAEELGDCGHVTSQDSQSWTSKNYPTNYTVCETIIVPK 67
QY 40 GNEITLLFMSGIIHVSGRGFLASY-----SVIDKQDLITCLDTASNFLPEFSKYCPAGCLL 95
Db 68 GKRL-IIRLGLDLNIESKTCASDYLLFSSATDQYDLITCLERGSHYFEKYKFCPCAGCD 126
QY 96 PFAEISGTIPHGVRDSSPLCMAGVHGVNTLGGQISVVISKIGIPIYESSLANNTSVV 155
Db 127 IAGDISGNTKGYRDTSLCKAAIHAGIITDELGHINLLQSKGISHYEGLLANGVLSRH 186
QY 156 GHLSLSTLFTFTSGCYCTLGMSVADPQITASSVLEWTDHTGQENSWKPKARLKKPG 215
Db 187 GSLSEKRLFTT-----198
QY 216 PPWAFATDEVQWLQIDLNKKTKITITGSTVMVHNYVSAYRILYSDDGQKWTYVRE 275
Db 199 -----198
QY 276 PGVEQDKIFQGNKYHODVRNLFPIIARFIRVNPTQWQKIAMKMLLGCQFIPKGR 335

unpublished
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 1388)
Shibata,T.
Direct Submission
Submitted (08-APR-1994) Tatsuhiko Shibata, National Cancer Center
Reserch Institute, Pathology; 1-1 Tsukiji 5 chome chuoku, Tokyo,
Tokyo 104, Japan (Tel:03-3542-2511(ex.4208), Fax:03-3248-2737)
On Mar 10, 1995 this sequence version replaced gi:474985.
Submitted (08-Apr-1994) to DDBJ by:
Tatsuhiko Shibata
National Cancer Center Reserch Institute
1-1 Tsukiji 5 chome
Chuoku, Tokyo 104
Japan
Phone: 03-3542-2511 x4208
Fax: 03-3248-2737.

FEATURES
source
Location/Qualifiers
1..1388
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="L121"
<1..1097
/codon_start=3
/product="unknown"
/protein_id="BAA18909.1"
/db_xref="GI:70441"
/translation="DGCCTHVLGPESTLTSINYPQVNSTVCWEIRVKMGSRVRI
KFGDPDIEDSDSCHFNRIYRNGIGVSRTEIGKYCGLGLOINHSIESKNGEITLIFMS
GTHVSRCFLASVYDKORNYLFGHCIOFFWNLSVSTAPAGCLLPFAEISGTIPH
GYRDSPLCMAGVAGVSVNTLGGQISVVISKGPYVESLIANNVTSVVGHLLIQVFP
TFKTSGCCVTGLMESGDRGSSNITVLEWTDHTGQENSKPKKSONKWTALGAF
AYDEQWQIDLNKKEKTKITGTTGGTGTGCTGGGGTGTGCAAAATCAACCAATTCA
ODKIFQGNRIIRIIRNRNLEPPIIARLLE"

BASE COUNT 432 a 289 c 302 g 364 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 1,59e-108 Length: 1388
Score: 1712.50 Matches: 374
Percent Similarity: 91.45% Conservative: 11
Best Local Similarity: 88.84% Mismatches: 19
Query Match: 49.35% Indels: 23
DB: 9 Gaps: 4

US-10-060-830-1114 (1-653) x HUMCUB1 (1-1388)

Qy 1 MetProLeuPheLeuLeuLeuValLeu-----LeuLeuLeuLeu 15
::: ||| :::::::::::::: |||
Db 148 TTGACATTGAAGATTCTGATTCCTGACCTTAATTAATTACTTGAGAAATTTATAATGGAATTG 207

Qy 16 GluAspAlaGlyAlaGlnGlnGlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSer 35
||| ||| :::::::::::::::
Db 208 GAGTCAGCAGA-ACTGAAATAGGCAAAATACTGTGGTCTGGGGTGTGCAAAATCAACCAATTCA 266

Qy 36 IleGluSerLysGlyAsnGluLeuThrLeuLeuPheMetSerGlyLeuHisValSerGly 55
|||||ATCAAAAGCAATGAATGACATTCCTGCTCATGAGTGAATCCAGTTCTGGA 326

Qy 56 ArgGlyPheLeuAlaSerTyrSerValIleAspLysGln-AspLeuIleThrCysLeuAs 75
|||||
Db 327 CGCGGATTTTGGCTCATACTCTGTATAGATAAACAAGATCTAATTAATTGTTGGA 386

Qy 75 pThrAlaSerAsn-PheLeuGluProGluPheSerLysTyrCys-ProAlaGlyCysLeu 94
|||||
Db 387 CACTGCATCCAAATTTTGGAACTCAGTTCAAGTACTGCCCGAGCTGGTGTCTG 446

Qy 95 LeuProPheAlaGluLeuSerGlyThrIleProHisGlyTyrArgAspSerProLeu 114
|||||
Db 447 CTTCCTTTTGTGAGATATCTGGAACAAATTCCTCATGATGATAGATTCCTCGCCATTG 506

Qy 115 CysMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerVal 134
|||||

Db 507 TGCATGGCTGGTGTGCATGCAGGAGTAGTGTCAACACACGTTGGCGCGCCAAATCAGTGT 566

Qy 135 ValIleSerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThrSerVal 154
|||||

Db 567 GTAATTAGTAAGGATTATCCCTTATTATGAAGTCTCTTGGCTAAACAACGTCACATCTGG 626

Qy 155 ValGlyHisLeuSerThrSerLeu---PheThrPheLysThrSerGlyCysTyrGlyThr 173
|||||::: |||||||

Db 627 GTGGGACACTTGATACTACAGTCTTTTACATTTAAGCAAGTGGATGTTATGGAACA 686

Qy 174 LeuGlyMetGluSerGly-ValIleAlaAspProGlnIleThrAlaSerValLeuG 193
|||||::: |||||||

Db 687 CTGGGATGGAGTCTGGTGTGATCGCGGATCCTCAATAACAGCATCACT-GTGCTGGA 745

Qy 193 uThrThrAspHisThrGlyGlnGluAsnSerTyrLys-ProLysLysAlaArgLeuLysL 213
|||||

Db 746 GTGGACTGACCACACAGGCAAGAACAGTGGAAACCCAAAAAAGGCGAGCTGAAA 805

Qy 213 ysProGlyProProTyrAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspL 233
|||||

Db 806 AACCTGGACCGCTTGGGT-GCTTTTGGCCACTGATGATACAGTGTTCACAAATAGATT 864

Qy 233 euAsnLysGluLysLysIleThrGlyIleThrGlySerThrMetValGluHis 253
|||||

Db 865 TGAATAAGGAAAAAGAAATAACAGGCATTATAACCACTGGATCCACCATGGT-GAGCACA 923

Qy 253 sTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysThrThrValT 273
|||||

Db 924 ATTACTATGTGTCTGCTCAGAAATCTGTACAGTGTATGCGCAGAAATGACGTGTGT 983

Qy 273 yArgGluProGly-ValGluGlnAspLysIlePheGlnGlyAsn--LysAspTyrHisG 292
|||||

Db 984 ACAGAGAGCTGTGCTGGGCAAGATAGATATTTCAGGAAACAAAGAAATATACACC 1043

Qy 292 InAsp-ValArgAsnAsnPheLeuProProIleIleAlaArgPheLeuArgValAsnPro 311
|||||

Db 1044 AGGATGGTGCCTAAATAACTTTTGGCCACCAATTTATTGTCAGG-TTTATTAGAGTGAATCCT 1102

Qy 312 ThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheLeuPro 331
|||||

Db 1103 ACCCAATGGCAGCAGAAATGCCATGAAATGAGCTGCTCGGATGTCAGTTTATTCCT 1162

Qy 332 LysGlyArgProProLysLeuThrGlnProProProProArgAsnSerAsnAspLeuLys 351
|||||

Db 1163 AAAGTCTGCTCCAAACTTACTCAACCTCCACCTCTCGGAACAGCAATGACCTCAA- 1221

Qy 352 AsnThrThrAlaProProLys---IleAlaLysGlyArgAlaProLysPheThrGlnPro 370
|||||

Db 1222 AACACTACAGCCCTCCAAAAATTTAGCCAAAGGTGTCGCCCAAAATTTACGCAACCA 1281

Qy 371 LeuGlnProArgSerSerAsnGluPheProAlaGlnThrGluGlnThrThrAlaSerPro 390
|||||

Db 1282 CTCAACCTCGCAGTAGCAATGATTTCTTCGCACAGACAGACAAACAACCACTGCCAGTCT 1341

Qy 391 AspIleArgAsnThrThrVal---ThrProAsnValThrLysAsp 404
|||||

Db 1342 GATATCAGAAATTAACCTACCGTAAACTCCAAATGTAACCAAGAT 1386

RESULT 9
AC013497 192247 bp DNA linear HTG 26-MAR-2000
LOCUS Homo sapiens chromosome 3 clone RP11-319J24 map 3, WORKING DRAFT
DEFINITION SEQUENCE, 11 unordered pieces.
AC013497
ACCESSION AC013497
VERSION AC013497.4 GI:7329299
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 192247)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-319J24

QY	249	MVEHNYVSAIRILYSDCGOKVTVYREPVEGQOKIPQGNKDYHQDVRRNLFPLPIIARFTR	308
DB	334	QSNFNEYVKSFMVNFNNNSKWKTYRGIVYNNNEKVFQGNSEFRDPQNNFIPIIVARYVR	393
QY	309	VNPTQHQOKTAKMKWELLGCOFIPKGRPPKLTQPPPPRSND---	364
DB	394	VVPQTQHQRIALKEVLGCO-----ITO-----GNDSLVWRKTSQSTSVSTKKED	438
QY	365	PKFTQLOPRSSNEFFPAQTEOTTASTPIDRNTVTTPNVTOKDALAAVLVPVLVWMLTTLIL	424
DB	439	ETITRI-----PSEET--STGINITT-----VAIPVLVLVLFVFGMGIFA	478
QY	425	ILVCAMHWRNRKKTKGTGYDLPLYW----DRAGWNKMGKOLFPAKAVDHEETP--VRYSSSE	479
DB	479	AF-----RKKKKKGS--PYGSAEAKTDCWKQIKY-----PFARHQSAAE	515
QY	480	--VNHLSPREVT---TVLQNDASAEYAQPLVGGIVGTLHORSTFKP---ESGKEAGYADLDP	532
DB	516	FTISYDNEKEMTOKLDLITSDMADYOOPLMIGTGTVRKGSTFRPMDTDAEEAG-VSTDA	574
QY	533	---YNSPGQVYHAYAEPLIPITGEYATPILMDMSGHPTTSVQGPSTSTSTFKATGNOP---	586
DB	575	GGHYDCFORAGRHEYALPLAPPEYATPIV---ERHVLRAHTFSAQSGYRVGPGQPGHK	631
QY	587	-----PPL--VGTYNTLLSRTDSCSSAAQAOYDTPKA	615
DB	632	HSLSGGGFSVAVGADGDGYORPHSAOAPDRGYDRPKA	670

```

RESULT 2
US-10-003-132-4
; Sequence 4, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG 2CUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-003-132-4

Query Match      12.0%; Score 417; DB 9; Length 503;
Best Local Similarity 22.4%; Pred. No. 3.9e-26;
Matches 148; Conservative 70; Mismatches 147; Indels. 296; Gaps 19;

```

Query Match	12.0%	Score 417;	DB 9;	Length 503;
Best Local Similarity	22.4%;	Pred. No. 3.9e-26;		
Matches 148;	Conservative 70;	Mismatches 147;	Indels. 296;	Gaps 19;
Qy	13	LLLEDAGAQOQKYCGGLGLQMNHSTESKNETTLFLFMGSHIHVSGRGFLASVVDKQDLIT	72	
Db	90	LLFSATDQIGPYCG-SWAVPKELRLNSNEVTLFKSGHISGKGLFLTYASSDHPDLIT	148	
Qy	73	CLDTAFNLEFPFKYCPAGCLLPFAISGTPHGYRDSPLCMAGHVAGVSNLTGGQI	132	
Db	149	CLERGSHYFEKYSKFCPAGCRDIAGDISGNTKDGRYDTSLLCKAAIHAGIITDELGGHI	208	
Qy	133	SVLISKGIPITYESSLANNVTSVGCHLSTLFTFTSGCYGTLGMSGVIAOPQITASSVL	192	
Db	209	NLLAQKGISHYEGLLANGVLSRHGSLSEKRFLETT-----	243	
Qy	193	EWTDDTGQENSMWPKKARLKKPGPPWAAAFATDEYQWLQIDLNKEKKITGITTGSTWVEH	252	
Db	244	-----	243	
Qy	253	NYVVSAYRILYSDGQKWTVYREFCVGEQDKIFQGNKDYHQDVRRNFLPPIIAREFVRNPT	312	

```

Db 244 -----PGM----- 246
QY 313 QWQOKIAMKMLLGCQIFPKGRPPKLTQPPPPRNSDLKNTTAPPKIAKRAPKFTQPLQ 372
Db 247 ----- 246
QY 373 PRSNEPPAQTEQTASPDIRNTVTNVTQKVALAALVVPVLVWLTTLILILVCANHW 432
Db 247 -----NIT-TVAIPSVI-FIALLLTGMGIFAIC----- 272
QY 433 RNKKKTECTYDLPY----WDRAGWKGKMQFLPAKAVDHEETP-VRYSSSE--VNHLS 485
Db 273 --RRKKKKGN---PYVSADAQKTCWKQIKY-----PFARHOSTEFTTISYDNE 315
QY 486 REVT---TVLOADSAEYAOQLVGGVIGTGHORSTFKP-----EE---GREAGYADLD 533
Db 316 KEMTQKLDLITSMDADYQQLMGTGTIVARKGSTFRPMDTDTTEVRVNTIEASGH---Y 370
QY 534 NSPQGEVYHAYAEPLPTGPEYATPII-----MDMSGHPTTSVGOQPSTSTFKATG 583
Db 371 DCPHRPGHEYALPLTHSEPEYATPIVERHLLRAHTFSTQSGYRVPG-PRPTHKSHSSG 429
QY 584 NQPPPLVGTINTLLSRDSCSSAQAOYDTPKACKPGLPADELVYQVPOSTQEVSGAGRD 643
Db 430 GFPPATGATQVESYQRPASPAPKPVGGYDKP-AASSFILDSRD-----PASQSOMTSGGDD 482
QY 644 G 644
Db 483 G 483

RESULT 3
US-09-823-038A-51
; Sequence 51, Application US/09823038A
; Patent No. US20020058335A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevin
; APPLICANT: Onrust, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Murison, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; FILE REFERENCE: 11000.1037c3
; CURRENT APPLICATION NUMBER: US/09/823,038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Mouse
US-09-823-038A-51

```

[illegible]

```

QY 137 SerLysGlyIleProTyrTyrGluSerSerLeuAlaAsnValThrSerValValGly 156
Db 537 AGTAAGGTATTCCTTATATCAAGATCTTGGCTACACGTCACATCTGTGGTGGGA 478
QY 157 HisLeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMet 176
Db 477 CACTTATCTACAAGTCTTTTACATTTAAGCAAGTGGTGTATGAACACTGGGGATG 418
QY 177 GluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluTrpThrAsp 196
Db 417 GAGTCTGTGTGATCGCGATCCTCAATAACAGCATCATCTGTCTGAGGGACTGAC 358
QY 197 HisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLysLysProGlyPro 216
Db 357 CACACAGGCAGAGACAGACAGTGGAAACCCAAAGCCAGCGCTCAAAAACCTGGACCC 298
QY 217 ProTfPAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLysGlu 236
Db 297 CTTGGCTGCTGCTTTTGGCCACTGATCAATACCAGTGGTTACAAATAGATTGAATAAGGAA 238
QY 237 LysLysIleThrGlyIleIleThrGlySerThrMetValGluHisAsnTyrTrpVal 256
Db 237 AAGAAATTAACAGGCATTATAACCACTGGAAATCACCATGGTGGAGCACAAATTAATGTG 178
QY 257 SerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTyrThrValTyrArgGluPro 276
Db 177 TCTGCCCTACAGAAATCTCTGTACAGTGNATGGCAGAAATGAGTGTGTACAGAGAGCT 118
QY 277 GlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspValArgAsn 296
Db 117 GGTGTGGAGCAAGATAAGATATTTCAAGCAACAAAGATTTATCACCAGGATGTGCGTAAT 58
QY 297 AspPheLeuProProlIleAlaArgPheIleArgValAsnProThrGlnTrpGln 315
Db 57 AACTTTTGGCCCAATATTATGACGCTTTTATTAGAGTGAATCTTACCACCAATGGCAG 1

RESULT 9
BB179946 643 bp mRNA linear EST 19-OCT-2001
LOCUS BB179946 RIKEN full-length enriched, adult male hypothalamus Mus
DEFINITION musculus cDNA clone A230081I02 3', mRNA sequence.
ACCESSION BB179946
VERSION BB179946.2 GI:162669978
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 643)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Jun 29, 2000 this sequence version replaced gi:8839029.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanadawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new

```

```

genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
FEATURES             Location/Qualifiers
     source            1..643
     organism="Mus musculus"
     db_xref="taxon:10090"
     clone_lib="RIKEN full-length enriched, adult male
     hypothalamus"
     sex="male"
     tissue_type="hypothalamus"
     dev_stage="adult"
     lab_host="DH10B"
     note="Site 1: Sali; Site 2: BamHI; cDNA library was
     prepared and sequenced in Mouse Genome Encyclopedia
     Project of Genome Exploration Research Group in Riken
     Genomic Sciences Center and Genome Science Laboratory in
     RIKEN. Division of Experimental Animal Research in Riken
     contributed to prepare mouse tissues. 1st strand cDNA was
     primed with a primer [5'
     GAGAGAGAGAAGGATCCAGAGCTCTTTTATTTTNN 3'], cDNA was
     prepared by using trehalose thermo-activated reverse
     transcriptase and subsequently enriched for full-length by
     cap-trapper. cDNA went through one round of normalization
     to Rot = 20.0 and subtraction to Rot = 459.0. Second
     strand cDNA was prepared with the primer adapter of
     sequence [5' GAGAGAGAGATTCGATTAATAATTAATCCCTCCCTCC
     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
     modified pBluescript KS(+) after bulk excision from Lambda
     FLC I."
BASE COUNT 168 a 130 c 171 g 174 t
ORIGIN
Alignment Scores:
Pred. No.: 1,38e-91 Length: 643
Score: 1023.00 Matches: 188
Percent Similarity: 94.31% Conservative: 11
Best Local Similarity: 89.10% Mismatches: 12
Query Match: 29.48% Indels: 0
DB: 10 Gaps: 0
US-10-060-830-1114 (1-653) x BB179946 (1-643)
QY 23 GlyLysTyrCysGlyLeuGlyLeuGlnMetAsnHisSerIleGluSerLysGlyAsnGlu 42
Db 11 GCCAATACTGTGGTCTGGTGTTCACAAATGAATCAATCAATTCAGTCCCAAGGACGTGA 70
QY 43 IleThrLeuPheMetSerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyr 62
Db 71 GTCACAGTCTGTTCATGATGAGTGAACCCATGCTGCTGGCGAGGATTTTGGCTTCATAC 130

```

Qy 63 SerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSerAsnPhelLeuGlu 82
 Db 131 TCAGTTATAGATAAGAAGATTAACTACTGTTTGGATGATCTAAATTTTGGAA 190
 Qy 83 ProGluPheSerLysTyrcysProAlaGlyCysLeuLeuProPheAlaGluIleSerGly 102
 Db 191 CCAGAGTTTCAGTAAGTACTGCCAGCTGGCTGCTCTTTTGGCTTTTGGCTGAATATCTGGA 250
 Qy 103 ThrIleProHisGlyTyrcysAspSerSerProLeuCysMetAlaGlyValHisAlaGly 122
 Db 251 ACAATCTCATGATACAGAGATCTTCCACATTTGTATGGCTGGATCCATCCAGCA 310
 Qy 123 ValValSerAsnThrLeuGlyGlnIleSerValIleSerLysGlyIleProTyr 142
 Db 311 GTAGTGTCAACAGCTGCTGGGTGGCCAAATCAGCATTTGTAGTAAAGGACCCCATAT 370
 Qy 143 TyrGluSerSerLeuAlaAsnValThrSerValValGlyHisLeuSerThrSerLeu 162
 Db 371 TATGAAGCTCTTTGGCCCAACAATCTCACTTCCACGGTGGGATCTTATCTGCAAGTCTG 430
 Qy 163 PheThrPheLysThrSerGlyCysTyrcysLeuGlyMetGluSerGlyValIleAla 182
 Db 431 TTTACATTTAGCAAGTGGTGTCTATGGACTCTGGGATGGAGTCTGGTGTGATGCC 490
 Qy 183 AspProGlnIleThrAlaSerSerValLeuGluTrpThrAspHisThrGlyGlnGluAsn 202
 Db 491 GATCCCCAGATAACAGCATCTGCTGCACCTGGAGTGGACTGACCCACATGGGGCAGGAGAAC 550
 Qy 203 SerTrpLysProLysAlaArgLeuLysLysProGlyProProTrpAlaAlaPheAla 222
 Db 551 AGCTGGACAGGGAGAGCCAGGCTGAGAAACCCGGGCCCTCCCTGGGCTGCTTTTGGC 610
 Qy 223 ThrAspGluTyrcysGlnTrpLeuGlnIleAspLeu 233
 Db 611 ACTGATGACATCAGTGGCTGCAGATAGACCTT 643

RESULT 10
 AW383889
 LOCUS
 DEFINITION QV3-HT0364-241199-024-B01 HT0364 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW383889
 VERSION AW383889.1 GI:6888548
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 664)
 HCGP <http://www.ludwig.org.br/ORESTES>.
 TITLE The FAPESP/LICR Human Cancer Genome Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?c1=QV3&t2=QV3-HT0364-241199-024-b01&t3=1999-11-24&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 53
 High quality sequence stop: 623.
 Location/Qualifiers
 1. 664
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HT0364"
 /dev_stage="Adult"

FEATURES
 source

/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 185 a 130 c 157 g 192 t
 ORIGIN
 Alignment Scores: 2.27e-90 Length: 664
 Pred. No.: 1011.00 Matches: 193
 Score: 96.53% Conservative: 2
 Percent Similarity: 95.54% Mismatches: 7
 Best Local Similarity: 29.14% Indels: 0
 Query Match: 10 Gaps: 0
 DB:
 US-10-060-830-1114 (1-653) x AW383889 (1-664)
 Qy 36 IleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHisValSerGly 55
 Db 53 ATTGAATCAAAAGCAATGAATCACATAGCTGTTTCATGAGTGAATCCATGTTCTGGA 112
 Qy 56 ArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAsp 75
 Db 113 CGCGGATTTTGGCCTCATCTCTGTATAGATAAACAAGATCTAATTAATGTTGTTGGAC 172
 Qy 76 ThrAlaSerAsnPhelLeuGluProGluPheSerLysTyrcysProAlaGlyCysLeuLeu 95
 Db 173 ACTGATCCCAATTTTGGAACTGAGTTCAGTAACTAGTCCCGGCTGGTGTCTGCTT 232
 Qy 96 ProPheAlaGluIleSerGlyThrIleProHisGlyTyrcysAspSerSerProLeuCys 115
 Db 233 CCTTTGCTGAGATATCTGGAACAATTCCTCATGATATAGAGATTCCTCGCCATTTGTC 292
 Qy 116 MetAlaGlyValHisAlaGlyValSerAsnThrLeuGlyGlyGlnIleSerValVal 135
 Db 293 ATGGCTGTGTGTCATGCGAGGAGTAGTCAACACAGCTTTGGCGGCCAAATCACTGTTGTA 352
 Qy 136 IleSerLysGlyIleProTyrTyrcysLeuLeuAlaAsnValThrSerValVal 155
 Db 353 ATTAGTAAGGTATTTCCCTATTATGAAAGTCTTTGGCTAACACGCTCACATCTGTGTG 412
 Qy 156 GlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrcysLeuGly 175
 Db 413 GGACACTTATCTACAAGTCTTTTACATTTAAGACAAGTGGATGTTATGGAACACTGGGG 472
 Qy 176 MetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluTrpThr 195
 Db 473 ATGGAGTCTGTGTGATCGCGGATCTCTCAAAATACAGCATCATCTGCTGGAGTGGACT 532
 Qy 196 AspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLysLysProGly 215
 Db 533 GACCCACAGGCCAGAGAACAGTGGAAACCCCAAAAGCCAGGCTGAAAAAACCTGGA 592
 Qy 216 ProProTrpAlaAlaPheAlaThrAspGluTyrcysGlnTrpLeuGlnIleAspLeuAsnLys 235
 Db 593 CGCGCTGGGCTGCTTTTGGCCACTGATGAATACAGAGTGGTTCCTCCCATAGAATTTGATTACG 652
 Qy 236 GluLys 237
 Db 653 GAAAGA 658
 RESULT 11
 BM781107
 LOCUS
 DEFINITION MLN1_4_E04.g1_A005 Mesenteric lymph node (MLN1) Equus caballus cDNA
 , mRNA sequence.
 ACCESSION BM781107
 VERSION BM781107.1 GI:19129339
 KEYWORDS EST.
 SOURCE horse.

Df 363 GTGGTGGACACATTTATCTACAAGTCCTTTTACATTAAAGACCAGCGGTGTTATGGRACA 422

Oy 174 LeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerValLeuGlu 193
|||||
Df 423 CTGGGATGGAATCTGGTGCTGATGCACAGATCACGGGCTCCTCTCTGCTGGAG 482

Oy 194 TrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLysLys 213
|||||
Df 483 TGGACTGACCACACAGGCCAAGAACAAGTTGGAAACCCCGANAGGCCGCTGAAGAAG 542

Oy 214 ProGlyProTrpAlaAlaPheAlaThrAspGluTrpGlnTrpLeuGlnIleAsp 232
|||||
Df 543 CCAGAGCTCTCTGGGCTGCACCTGGCCACTGACGAGTACCAGTGGTTACNAGTCGAC 599

RESULT 12
BB657747 LOCUS BB657747 608 bp mRNA linear EST 26-OCT-2001
DEFINITION Musculus cdna clone D330015M16 5', mRNA sequence.
ACCESSION BB657747
KEYWORDS house mouse.
SOURCE EST.
ORGANISM Mus musculus
REFERENCE Eukaryote; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
M., Koye,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Sasaki,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
D., Shibata,K., Shingawa,A., Shiraki,T., Sobabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T.,
Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
COMMENT Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: //genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukuishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamana,I., Aizawa,
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.
Location/Qualifiers
1..608
/organism="Mus musculus"

```
/db_xref="taxon:10090"
/clone="D230015M16"
/clone_lib="RIKEN full-length enriched, 12 days embryo
eyeball"
/tissue_type="eyeball"
/lab_host="DH10B"
/dev_stage="12 days embryo"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGCGCGCGCACTCGAGTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5',
GAGAGAGAGATTCGAGTTAATTAAATTCCTCCCGCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 162 a 126 c 159 g 161 t
ORIGIN

Alignment Scores:
Pred. No.: 4.92e-87 Length: 608
Score: 977.00 Matches: 180
Percent Similarity: 94.06% Conservativeness: 10
Best Local Similarity: 89.11% Mismatches: 12
Query Match: 28.46% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-1114 (1-653) x BB657747 (1-608)
QY 49 SerGlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGln 68
Db 1 AGTGAACCCATGCTGCTGGCGGAGGATTTTGGCTTCATCTACTCAGTTATAGATAAGA 60
QY 69 AspLeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyr 88
Db 61 GATTTAATCAGCTGTTGGATACGTATCTAATTTTGGAAACCCAGAGTTCAGTAAGTAC 120
QY 89 CysProAlaGlyCysLeuLeuPropheAlaGluIleSerGlyThrIleProHisGlyTyr 108
Db 121 TGCCCGAGCTGGCTGCTTTTGGCTTTGCTGAAATATCTGGAAACAATTCCTCATGATAC 180
QY 109 ArgAspSerSerProLeuCysMetAlaGlyValHisAlaGlyValValSerAsnThrLeu 128
Db 181 AGAGATTCTTCACCATTTGTATGGCTGGAATCCATGCAGGAGTAGTGCAAAACGTGCTG 240
QY 129 GlyGlyGlnIleSerValValIleSerLysGlyIleProTyrTyrGluSerSerLeuAla 148
Db 241 GGTGGCCCAATCAGCATTTGATTAACAAGGAGGCCCAATATTAAGAGCTCTTTGGCC 300
QY 149 AsnAsnValThrSerValValGlyHisLeuSerThrSerLeuPheThrPheLysThrSer 168
Db 301 AACAATGTCACATCCACGGTGGGATCTATCTGCAAGTCTGTTACATTTAAGACAAGT 360
QY 169 GlyCysTyrGlyThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAla 188
Db 361 GGTTCATATGGGACTCTGGGGATGGAGTCTGTGTGATTGCCGATCCCAAGATAACAGCA 420
QY 189 SerSerValLeuThrThrAspHisThrGlyGlnGluAsnSerTrpLysLys 208
Db 421 TGTCTGCTACTGGAGTGGAGTACCACATGTGGGCGAGAGAACAGCTGGACGGGAGAG 480
QY 209 AlaArgLeuLysLysProGlyProProTyrAlaAlaPheAlaThrAspGluTyrGlnTrp 228
Db 481 GCCAGCTGAGAAACCCGGCCCTCCCTGGCTGCTTTTCCCACTGATGAGCATCATGG 540
QY 229 LeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleThrThrGlySerThr 248
Db 229 LeuGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleThrThrGlySerThr 248
```

```
Db 541 CTGCAGATAGACCTTAACAAGGAGAAGACGATACAGGCTATCGTAACCACTGGGTCTACC 600
QY 249 MetVal 250
||||:
Db 601 ATGATA 606

RESULT 13
AV603144/c
LOCUS
DEFINITION AV603144 Bos taurus kidney fetus Bos taurus cDNA clone E1K1014D09
5', mRNA sequence.
ACCESSION AV603144
VERSION AV603144.1 GI:9725470
KEYWORDS EST.
SOURCE
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 575)
AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
TITLE poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE 21570554
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@coea.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
FEATURES
source
location/Qualifiers
1..575
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="E1K1014D09"
/clone_lib="Bos taurus kidney fetus"
/tissue_type="kidney"
/dev_stage="fetus"
/lab_host="DH10B"
/note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 131 a 137 c 129 g 178 t
ORIGIN

Alignment Scores:
Pred. No.: 9.05e-87 Length: 575
Score: 974.00 Matches: 176
Percent Similarity: 97.87% Conservativeness: 8
Best Local Similarity: 93.62% Mismatches: 4
Query Match: 28.07% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-1114 (1-653) x AV603144 (1-575)
QY 174 LeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerValLeuGlu 193
Db 575 CTAGGCATGGAATCTGGGTGATCGCTGATGCTCAATACCAAGCATCATCTGTGTGGAG 516
QY 194 TrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLysLys 213
Db 515 TGACCGGACACACAGACAGACAGACAGACAGCTGGAAACCTGAAAGCCAGCGTGAAGAAA 456
QY 214 ProGlyProProTyrAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeu 233
Db 455 CTGGACCTCTCTGGGCTGCTTTTGGCAGTGAATATCATGCTGTTGCAAGTAGATCTG 396
QY 234 AsnLysGluLysLysIleThrGlyIleThrThrGlySerThrMetValGluHisAsn 253
```

```

|||||
395 AATAAGGAAAGAAAGATCACAGGCAATGTGACTACTGCTCCACCATGGTGGACACAAT 336
Qy 254 TyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTyrThrValTyr 273
Db 335 TACTACGTGCTCAGCTCAAGATCTATACAGTCTGCTATGGCAGAGGTGGACTGTGTAC 276
Qy 274 ArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyrHisGlnAsp 293
Db 275 AGAGAGCTGTGTGGACGAGATGATATTTCAAGAAACAAGATTATCACCAGGAT 216
Qy 294 ValArgAsnAsnPheLeuProPheIleAlaArgPheIleAlaArgValAsnProThrGln 313
Db 215 GTGCCTAATAACTTTTCCACCAATTTATTCAGTGAATCTTATTTATTCCTAAAGGT 156
Qy 314 TrpGlnGlnLysIleAlaMetLysMetGluLeuGlnGlyCysGlnPheIleProLysGly 333
Db 155 TGGCAGCAGAAATTTGCCATGAAATGGAAGTGTGCTGATGCTCAATTTATTTCTTAAAGGT 96
Qy 334 ArgProProLysLeuThrGlnProProProProArgAsnSerAsnAspLeuLysAsnThr 353
Db 95 CGTCTCCAAACTTACTCAACCTCCGCTCCTCGAGACGAATGACCTCAGAAATACT 36
Qy 354 ThrAlaProProLysIleAlaLys 361
Db 35 ACAACCCCTCCAAATAAGCCAAA 12

RESULT 14
BF167024 912 bp mRNA linear EST 30-OCT-2000
LOCUS 601774167F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3994745 5',
DEFINITION mRNA sequence.
ACCESSION BF167024
VERSION BF167024.1 GI:11047286
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 912)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14W9213 row: e column: 18
High quality sequence stop: 651.
Location/Qualifiers
1. 912
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site1: SalI;
Site2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 233 a 265 c 256 g 158 t
ORIGIN

Alignment Scores:
Pred. No.: 1.75e-86 Length: 912

```

```

Score: 974.00 Matches: 209
Percent Similarity: 87.60% Conservative: 10
Best Local Similarity: 83.60% Mismatches: 27
Query Match: 28.07% Indels: 9
DB: 12 Gaps: 3

US-10-060-830-1114 (1-653) x BF167024 (1-912)

Qy 393 ArgAsnThrValThrProAsnValThrLysAspValAlaLeuAlaValLeuVal 412
Db 1 AAAACACAGCTGTAACTCCAGTGTAACCAAGATGCGCACTGGCTGCTGCTTCTGGTC 60
Qy 413 ProValLeuValMetValLeuThrThrLeuLeuLeuLeuValLysAlaTrpHisTrp 432
Db 61 CCTGTGCTGGTCATGGCCCTCACCACACTCATCTCTAGTGTGTGTGGCACTGG 120
Qy 433 ArgAsnArgLysLysLysThrGluGlyThrTyrAspLeuProTyrTrpAspArgAlaGly 452
Db 121 AGAAACAGGAAAGAAAGAACTGAAGCGCTATGATTTACCCCACTGGGATCGGCAGGT 180
Qy 453 TrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluThrPro 472
Db 181 TGGTGGAAAGGAATGAAGCAGCTTCTCCCTGCCAAGTCGGTGGACCACGAGGAGGCCA 240
Qy 473 ValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrThrValLeu 492
Db 241 GTGGCTACAGCACTAGTGAAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGT 300
Qy 493 GlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyLeuValGlyThrLeuHis 512
Db 301 CAGGCCGACTCTGCAGATATATGCACAGCCCTCGTGGGAGGAATTTGTTGGCACACTCCAT 360
Qy 513 GlnArgSerThrPheLysProGluGlyLysGluAlaGlyTyrAlaAspLeuAspPro 532
Db 361 CAGAGATCCACCTTTAAACCTCAGGAAGGAAGGAGGAGGCTATGCACACCTCATCTCT 420
Qy 533 TyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuProLysThrGly 552
Db 421 TACAACTCTCCAGTACAGGAGTGTACCACGCTATGCTGAACCACTGCCCTTAACGGGG 480
Qy 553 ProGluTyrAlaThrProIleMetAspMetSerGlyHisProThrThrSerValGly 572
Db 481 CTTAGTAGCGCAAC-CCGATGCTATGGACATGTGAGGCGACCCACACCTCAGTTGTT 539
Qy 573 GlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProProLeuValGlyThr 592
Db 540 CTGGCTCC-ACATCACTTCAAACTGAGG---GACCCAGCTCCCGCTTAGTGGGAAT 595
Qy 593 TyrAsnThrLeuLeuSerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThr 612
Db 596 TACAACAT-CTTCTCTCCAGGACTGACAG-TGTCTCTCGGGCAGGCTCAGTATGACACC 653
Qy 613 ProLysAlaGlyLysProGlyLeuProAlaProAspGluLeuValTyrGlnValProGln 632
Db 654 GCAAAAGGTGGGAAGTCAGTGCTA-----CCCCAGAGAAACGGTGTACCACTGGCCAGAG 707
Qy 633 SerThrGlnGluValSerGlyAlaGlyArg 642
Db 708 ---ACCCAGGAGGTATCAGA-CCAGGAAGG 733

RESULT 15
BQ037529 628 bp mRNA linear EST 01-MAY-2002
LOCUS BQ037529
DEFINITION pgnc.p0006.i10 normalized chicken lymphoid cDNA library Gallus
gallus cDNA clone pgnc.p0006.i10 5' similar to ref|X059528.2
(XM_059528) endothelial and smooth muscle cell-derived
neurofilin-like protein [Homo sapiens], mRNA sequence.
ACCESSION BQ037529
VERSION BQ037529.2 GI:20382257
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 628)

Morgan, R.W. and Burnside, J.

Chicken lymphoid ESTs

Unpublished (2001)

On Mar 27, 2002 this sequence version replaced gi:19771069.

Contact: Joan Burnside

Molecular Endocrinology

University of Delaware

40 Townsend Hall, Newark, DE 19717, USA

Tel: 302 831-1345

Fax: 302-831-3411

Email: joan@udel.edu, www.chickest.udel.edu.

FEATURES

Source

1..628
Location/Qualifiers
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="pgn1c.pk006.i10"
/clone_lib="normalized chicken lymphoid cDNA library"
/sex="Male and Female"
/tissue.type="thymus, bursa, spleen, PBL, bone marrow"
/lab_host="E.coli BMDH10B"
/note="Vector: pCMVSPORT 6"

BASE COUNT 186 a 154 c 143 g 143 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 1..62e-86 Length: 628
Score: 972.00 Matches: 180
Percent Similarity: 91.25% Conservative: 10
Best Local Similarity: 86.54% Mismatches: 18
Query Match: 28.01% Indels: 0
DB: 14 Gaps: 0

US-10-060-830-1114 (1-653) x BQ037529 (1-628)

QY 393 ArgAsnThrValThrProAsnValThrLysAspValAlaLeuAlaValLeuVal 412
Db 1 AAAAACCACCACTGACTCCAAAGCGTAAACCAAGATGTGGCAGCAGTTCTGGTT 60
QY 413 ProValLeuValMetValLeuThrLeuLeuLeuValCysAlaTrpHisTrp 432
Db 61 CCAGTGTGTGTGATGGTCTTCACTACTCTCTATCTTATCTTGTGTGGCATTTGG 120
QY 433 ArgAsnArgLysLysLysThrGluGlyThrTyrAspLeuProTyrTrpAspArgAlaGly 452
Db 121 AGAACCGCAAGAAACAAACGAGGGCAGCTATGACCTACCTACTGGATCGTGCAGG 180
QY 453 TrpTrpLysGlyMetLysGlnPheLeuProAlaLysAlaValAspHisGluGlyThrPro 472
Db 181 TGGTGAAGGAATGAAGCAGTTTCTCCCTACGAAATCAGCAGAACATGAAGAACTCCT 240
QY 473 ValArgTyrSerSerSerGluValAsnHisLeuSerProArgGluValThrValLeu 492
Db 241 GTACGCTACAGCAGCAGTGAATTTGGCCGCTCAGACCAAGAGAAAGTTCCAAACAATGCTG 300
QY 493 GlnAlaAspSerAlaGluTyrAlaGlnProLeuValGlyGlyIleValGlyThrLeuHis 512
Db 301 CAGACAGAACTCGCAGAGTACGCTCAGCCCTCTGGTAGGGGCATAGTGGCAGCTCTTCAT 360
QY 513 GlnArgSerThrPheLysProGluGlyLysGluAlaGlyTyrAlaAspLeuAspPro 532
Db 361 CAGAGATCAACCTTCAACACAGAGGAGGAAAGAAAGAGCGAGTTTATGCTGATTTGGACCT 420
QY 533 TyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIleThrGly 552
Db 421 TACAATTACCTATACAGAAGTTTATCATGCCCTATGCTGAACCACTGCCTATAACTGGA 480
QY 553 ProGluTyrAlaThrProIleIleMetAspMetSerGlyHisProThrThrSerValGly 572
Db 481 CCAGATACGCAACTCCAAATAGTCATGGACATGTCCAGTTCATCCAGCACACCTCTTGGT 540

QY 573 GlnProSerThrSerThrPheLysAlaThrGlyAsnGlnProProLeuValGlyThr 592
Db 541 GTTCCTTCTATTTCACCTTCAAAGCAGCAGGGAATCAAGCTCCTTCACTGGTGGAACT 600
QY 593 TyrAsnThrLeuLeuSerArgThr 600
Db 601 TACAATAAGCTCTTGTCCAGGACA 624

Search completed: January 23, 2003, 15:16:39
Job time : 2326.88 secs

Alignment Scores:			
Pred. No.:	2.92e-79	Length:	640
Score:	897.00	Matches:	194
Percent Similarity:	91.98%	Conservative:	1
Best Local Similarity:	91.51%	Mismatches:	17
Query Watch:	25.85%	Indels:	6
DB:	4	Gaps:	0
US-10-060-830-1114 (1-653) x US-09-328-111-480 (1-640)			
Qy	299	LeuProProIlelleAlaAArgPheIleAArgValAsnProThrGlnTrpGlnGlnLysIle	318
Db	630	TTGCCACCAATATAGGCAGC-TTTATTAGAGT-AATCTTACCATTGGCAGCAGAAAT-	574
Qy	319	AlaMetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeu	338
Db	573	GCCATGAAAT-GAACTGCTCGGATGTCAGTTTATTCNTAAAGT-NGTCNTCCAAACHT	516
Qy	339	ThrGlnProProProAArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLys	358
Db	515	ACTCAACTCCACTCTTNGNAACAGCAATGNCCTCAAAACCACTACAGCCCTCCAAAA	456
Qy	359	IleAlaLysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGlu	378
Db	455	TA-GCCAAAGGTCTGCCCCCAAATTTACGCAACCACTCAACTTCGCCAGTAGCAATGAA	397
Qy	379	PheProAlaGlnThrGluGlnThrThrAlaSerProAspIleAArgAsnThrThrValThr	398
Db	396	TTTCTCTGCACAGACAGAAACAACACTGCCAGTCTCTGATATCAGAAATACTACCGTACT	337
Qy	399	ProAsnValThrLysAspValAlaLeuAlaAlaValLeuValProValLeuValMetVal	418
Db	336	CCAAATGTAACCAAGATGTAGCGCTGGCTGCGAGTTCTTGTCCCTGTCTGCTCATGGTC	277
Qy	419	LeuThrThrLeuIleLeuIleLeuValCysAlaTProHisTrpArgAsnArgLysLysLys	438
Db	276	CTCACTACTCTCATTCATTTAGTGTGTGTGTGGCACTGGAGAACAGAAAGAAAAA	217
Qy	439	ThrGluGlyThrTyrAspLeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLys	458
Db	216	ACTGAAGGCACCTATGACTTACCTTACTGGGGCCGGCAGGTTGGTGGAAAGGAATGAAG	157
Qy	459	GlnPheLeuProAlaLysAlaValAspHisGluThrProValArgTyrSerSerSer	478
Db	156	CAGTTTCTTCCTCCAAAGCAGTGGACCATGAGGAACCCCACTGCTGCTATTAGCAGCAGC	97
Qy	479	GluValAsnHisLeuSerProAArgGluValThrThrValLeuGlnAlaAspSerAlaGlu	498
Db	96	GAAGTTAATCACTGAGTCCAGAGAAAGTCACCAAGTCTCGAGCTGACTCTCGACAG	37
Qy	499	TyrAlaGlnProLeuValGlyGlyIleValGlyThr	510
Db	36	TATGCTCAGCACTGCTAGGAGAAATGTTGTGATC	1

```

RESULT 2
US-08-746-111-4
; Sequence 4, Application US/08746111
; Patent No. 6066778
; GENERAL INFORMATION:
; APPLICANT: Ginsburg, David
; APPLICANT: Cui, Jisong
; TITLE OF INVENTION: Compositions And Methods For Screening
; TITLE OF INVENTION: Compounds For Anticoagulant Activity
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

[illegible]


```

RESULT 4
US-08-658-340-1
; Sequence 1, Application US/08658340
; Patent No. 5910576
; Patent No. 5910576 5861489
; GENERAL INFORMATION:
; APPLICANT: Bertina, Rogier
; APPLICANT: Reitsma, Pieter
; TITLE OF INVENTION: A method for diagnosing an increased
; TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
; TITLE OF INVENTION: thrombosis and kit for use with the same.
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5910576 5861489el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,340
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/454,353
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 948-7400
; TELEFAX: (301) 948-9751
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-658-340-1

Alignment Scores:
Pred. No.: 4,23e-24 Length: 6909
Score: 353.50 Matches: 108
Percent Similarity: 44.20% Conservative: 33
Best Local Similarity: 33.86% Mismatches: 116
Query Match: 10.19% Indels: 62
DB: 2 Gaps: 11

US-10-060-830-1114 (1-653) x US-08-658-340-1 (1-6909)
QY 50 GlylleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAsp 69
Db 5862 GCCTTCAGAGTTCTCGGGTTA---CTGGAGCCCGAGATTAGCAAGATTAAACAATGGTGG 5918
QY 70 LeulleThrCysLeuAspThrAlaSerAsnPhleuGluProGluPheSerLysTyrCys 89
Db 5919 ATCTTATAATGCTTG-----GAGTGAGAAAACATTGC 5951
QY 90 ProAlaGlyCys-----LeuLeuProPheAlaGluLeSerGlyThrIleProHis 106
Db 5952 AGCAGAAATTCGCTCTAAACCTTGGATCCAGTGACATGCAAAAGGAAGT-----CAT 6005
QY 107 GlyTyrArgAspSerSerProLeuCysMethala-----Gly 118
Db 6006 AATCACAGGATCCAGACCCAGAGGTGCCAAACACTACTGAGTCTGCTATACACAGA 6065
QY 119 Val-HisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValIleSer-- 137
Db 119 Val-HisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValIleSer-- 137

```

```

Db 6066 GTTCTATGTAGCTTACAGTTCCAAACCCAGATCAACTGGCAGATCTTCAAAGGAACAGCAC 6125
QY 138 -LysGlyIleProTyrTyrGluSerSerLeuAlaAsnAsnValThrSerValValGlyHi 157
Db 6126 AAGGATGTGATGTATTTTAAT-----GGCAATTCAGATGCCTCTACAATAAAGA 6176
QY 157 s-----157
Db 6177 GAATCAGTTTGACCCACCTATTGTGCTAGATATATTAGGATCTCTCCAACTCAGAGCCTA 6236
QY 158 -----LeuSerThrSerLeuPheThrPheLeuThrSerGlyCysTyrGlyTh 173
Db 6237 TAACAGACTACCTTCGATTGCAAGTGTGAGGTAATAGGATGTTTCCACACC 6296
QY 173 rLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGl 193
Db 6297 CTGGGTATGGAAATAGGAAGATAGAAACAGCAATACACAGCTTCTTCGTTTAAAGAA 6356
QY 193 u-----TirThrAspHisThrGlyGlnGluAsnSerTirLysProLysLysAlaArgLe 211
Db 6357 ATCTTGGTGGGAGATTAC-----TGGGAACCTCTCCGTGCGCGTCT 6398
QY 211 uLysLysProGly-----ProProTirpalaalaphelaThrAspGluTyrGlnTirple 229
Db 6399 GAATGCCAGGACGCTGTGAATCCCTGGCAAGCCAAAGCAACAACAATAAGCAGTGGCT 6458
QY 229 uGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleIleThrGlySerThrMe 249
Db 6459 AGAAATGATCTACTCAAGATCAGAGATAGAGGCAATTAACACAGGCTGCAAGTC 6518
QY 249 tValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnly 269
Db 6519 TCTGTCTCTGAAATGTATGTAAGAGCTATACCATCCACTACAGTGACGAGGAGTGA 6578
QY 269 strpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAs 289
Db 6579 ATGGAACCATACAGGCTGAAATCCCTCATGCTGGCAACAAGATTTTGAAGGAATACTAA 6638
QY 289 pTyrHisGlnAspValArgAsnAsnPhleuLeuProIleIleAlaAtrpHeileArgVa 309
Db 6639 TACCAAGGACATGTGAAGAACTTTTCAACCCCAATCATTTCCAGTTTATCCGTGT 6698
QY 309 lAsnProThrGlnTirpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCys 327
Db 6699 CATTCCTAAACATGGAATCAAGTATTACACTTCGCTGGAACCTCTTTGGCTGT 6753
RESULT 5
US-08-746-111-26
; Sequence 26, Application US/08746111
; Patent No. 6066778
; GENERAL INFORMATION:
; APPLICANT: Ginsburg, David
; APPLICANT: Cui, Jisong
; TITLE OF INVENTION: Compositions And Methods For Screening
; TITLE OF INVENTION: Compounds For Anticoagulant Activity
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,111
; FILING DATE: 06-NOV-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

```

NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 6909 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-746-111-26

Alignment Scores:
Pred. No.: 4,23e-24 Length: 6909
Score: 353.50 Matches: 108
Percent Similarity: 44.20% Conservative: 33
Best Local Similarity: 33.86% Mismatches: 116
Query Match: 10.19% Indels: 62
DB: 3 Gaps: 11

US-10-060-830-1114 (1-653) x US-08-746-111-26 (1-6909)

QY 50 GlyIleHisValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAsp 69
Db 5862 GCCTCAGAGTTCTGGGTTA---CTGGAGCCAGATTAGCAAGATTAACATGTTGG 5918
QY 70 LeuIleThrCysLeuAspThrAlaSerAsnPheLeuGluProGluPheSerLysTyrCys 89
Db 5919 ATCTTATAATGCTTG-----GAGTGTAGAAAAAATGTC 5951
QY 90 ProIleGlyCys-----LeuLeuProPheAlaGluIleSerGlyThrIleProHis 106
Db 5952 AGCAGAAATTTGCTCTAAACCTTCGAGTCCAGGTGGACATGCAAAAGGAAGT-----CAT 6005
QY 107 GlyTyrArgAspSerProLeuCysMetaIa-----Gly 118
Db 6006 AATCAGAGGATCCAGACCCAGTGCACAACTACTGAGTCTGCTATACACAGA 6065
QY 119 Val-HisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValValIleSer-- 137
Db 6066 GTTCTATGTAGTTACAGTTCCACACAGATCAACTGGCAGATCTCAAAAGGAACAGCAC 6125
QY 138 -LysGlyIleProTyrThrGluSerSerLeuAlaAsnValThrSerValValGlyHi 157
Db 6126 AAGGAATGTGATGATTTTAAAT-----GGCAATTTCAGATGCCTCTACAATAAAGA 6176
QY 157 s----- 157
Db 6177 GAATCAGTTGACCCACCTATTGTGGCTAGATATATTAGGATCTCTCAACTCGAGCCTA 6236
QY 158 -----LeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGlyTh 173
Db 6237 TAACAGACCTACCTTCGATTGGAACTGCAAGGTTGTGAGTAAATGATGCTTCCACACC 6296
QY 173 rLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuG1 193
Db 6297 CTTGGGTATGAAATGGAAGATAGAAAACAAGCAATATCACAGCTTCTCGTTTAAGAA 6356
QY 193 u-----TrpThrAspHisThrGlyGlnGluAsnSerThrLysProLysLysAlaArgLe 211
Db 6357 ATCTTGGTGGGAGATTAC-----TGGGAACCTTCCTCGGCGCT 6398
QY 211 uLysLysProGly-----ProProTrpPalaAlaPheAlaThrAspGluTyrGlnTrpLe 229
Db 6399 GAATCCAGGACGCTGTGAATGCTGCGAGCCAGCAAGCAACAAACATAAGCAGTGGCT 6458
QY 229 uGlnIleAspLeuAsnLysGluLysLysIleThrGlyIleThrThrGlySerThrMe 249
Db 6459 AGAATTCATCTACTACAGATCAAGAGATAACGGCAATTATAACACAGGGCTGCAAGTC 6518

QY 249 tValGluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLy 269
Db 6519 TCTGTCTCTGAAATGATGTAAGAGCTATACCACTCCATCAGTACGAGGAGTGGGA 6578
QY 269 sTrpThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAs 289
Db 6579 ATGGAACACATACAGCTGGAATCCCTCCATGCTGGTGGACAGATTTTGAAGGAATACTAA 6638
QY 289 pTyrHisGlnAspValArgAsnAsnPheLeuProProIleAlaAlaArgPheIleArgVa 309
Db 6639 TACCAAGGACATGTGAAGAACTTTTCAACCCCACTCATTTCCAGGTTTATCCGTGT 6698
QY 309 lAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCys 327
Db 6699 CATTCCTAAACATGGAATCAAGATATTACACTTCGCTCGAACTCTTTGGCTGT 6753

RESULT 6
US-08-480-229C-11
Sequence 11, Application US/08480229C
Patent No. 5874562
GENERAL INFORMATION:
APPLICANT: Quettermous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,229C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1779
US-08-480-229C-11

Alignment Scores:
Pred. No.: 6.68e-25 Length: 1780
Score: 352.00 Matches: 75
Percent Similarity: 58.72% Conservative: 26
Best Local Similarity: 43.60% Mismatches: 67
Query Match: 10.14% Indels: 4
DB: 2 Gaps: 2

US-10-060-830-1114 (1-653) x US-08-480-229C-11 (1-1780)

:	INFORMATION FOR SEQ ID NO:	11:			
:	SEQUENCE CHARACTERISTICS:				
:	LENGTH:	1780 base pairs			
:	TYPE:	nucleic acid			
:	STRANDEDNESS:	single			
:	TOPOLOGY:	unknown			
:	MOLECULE TYPE:	CDNA			
:	FEATURES:				
:	NAME/KEY:	CDS			
:	LOCATION:	1..1779			
:	US-08-	659-235C-11			
<hr/>					
Alignment Scores:					
Pred.	No.:	6.68e-25	Length:	1780	
Score:		352.00	Matches:	75	
Percent Similarity:		58.72%	Conservative:	26	
Best Local Similarity:		43.60%	Mismatches:	67	
Query Match:		10.14%	Indels:	4	
DB:		2	Gaps:	2	
<hr/>					
US-10-060-830-1114 (1-653) x US-08-659-235C-11 (1-1780)					
Qy	158	LeuSerThrSerLeuPheThrPhelyThrSerGlyCysTyrGlyThrLeuGlyMetGlu	177		
Dd	1105	TTCGCAATGGAACTCTCGGTGTGAACATGTCGGGTGTTCTAGCCCTCTCGGGTATGAAA	1164		
Qy	178	SerGlyValIleAlaAspProGlnIleThrAlaSerServallLeuGluTrpThrAspHis	197		
Dd	1165	TCAGGCATATACAAGACTATCATCACTGCTCCAGCATCTTCAAGAAGCTCAACATG	1224		
Qy	198	ThrGlyGlnGluAsnSerTriplyProLysLysAlaArgLeuLysLysProGlyPro---	216		
Dd	1225	-----GACATGTTCCATTGGGAACCAGGAAGCTCGGCTGGACAAGCAAGCAAAGTG	1278		
Qy	217	---ProtipAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsnLys	235		
Dd	1279	AATGCTGGACCTCTGCCCAACATGACCAGTCACAAATGGTTACAGGGTGGATCTCTTGGT	1338		
Qy	236	GluLysLysIleThrGlyIleIleThrThrGlySerThrMetValGluHisAsnTyrTyr	255		
Dd	1339	CCAACCAAGTAGCTGGCATCATTAACAAGGAGCTAAAGATTGTGTCATGTACAGTTT	1398		
Qy	256	ValSerAlaTyrArgIleLeuTyrSerAspaspGlyGlnLysTrpThrValTyrArgGlu	275		
Dd	1399	GTTGGCTCTCAAACTGGCTTACAGCAATCATGGAGAACACTGGACTGTATACCAAGAT	1458		
Qy	276	ProGlyValGluGlnAspLysIlePheGlnGlvsnLysAspTyrHisGlnAspValarg	295		
Dd	1459	GAAGAACCAAGCAAMAGATAAGTTTTCCAGGGAATTTTCAGCAATGACACTCACAGAAA	1518		
Qy	296	AsnAsnPheLeuProIleAlaAArgPheIleArgValasnProThrGlnTrpGln	315		
Dd	1519	AATGTCATCGACCTCCCATCTATGACGACACACATAGATCCTTCCTGGTCTGGTAC	1578		
Qy	316	GlnLysIleAlaMetLysMetGluLeuGlyCys	327		
Dd	1579	GGGAGGATCACATTGGCGTCAGAGCTGCTGGGCTGC	1614		
<hr/>					
RESULT 8					
US-08-480-229C-9					
; Sequence 9, Application US/08480229C					
; Patent No. 5874562					
GENERAL INFORMATION:					
APPLICANT: Quattermous, Thomas					
APPLICANT: Hogan, Bridgid					
APPLICANT: Snodgrass, H. Ralph					
APPLICANT: Zupancic, Thomas J.					
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL					
NUMBER OF SEQUENCES: 29					
CORRESPONDENCE ADDRESS:					
ADDRESSEE: Pennie & Edmonds LLP					
STREET: 1155 Avenue of the Americas					

```

Db 1951 AATGTCATCGACCCCTCCCATCTATCCAGCATTCATAAGAAATCCCTTCTTGGTCTCGGTAT 2011
Qy 316 GlnLysIleAlaMetLysMetGluLeuLeuGlyCys 327
Db 2011 GGAAGGATCACTCTGGGTCAGAGCTGCTGGGCTGC 2046
:::||||| ::::: |||||
RESULT 9
US-08-659-235C-9
; Sequence 9, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Shodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,235C
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0034-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 619..2058
US-08-659-235C-9
Alignment Scores:
Pred. No.: 1,01e-24 Length: 2303
Score: 352.00 Matches: 75
Percent Similarity: 59.30% Conservative: 27
Best Local Similarity: 43.60% Mismatches: 66
Query Match: 10.14% Indels: 4
DB: 2 Gaps: 2
US-10-060-830-1114 (1-653) x US-08-659-235C-9 (1-2303)
Qy 158 LeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMetGlu 177
Db 1537 TTAAGAATGGAACCTCTTGGCTGTGAGCTCTCAGGCTGTTCAGAACCTTTGGGGATGAAA 1596
Qy 178 SerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluTrpThrAspHis 197
Db 1597 TCAGGGCATATACAGACTACACAGATCATCTGCTCCAGGCTTTCAGAACACTCAACATG 1656
Qy 198 ThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLysLysProGlyPro--- 216

```


;; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
;; TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
;; NUMBER OF SEQUENCES: 29
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pretty, Schroeder & Poplawski
;; STREET: 444 South Flower St., 19th Floor
;; CITY: Los Angeles
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 90071
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/162,402B
;; FILING DATE: 03-DEC-1993
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Amzel, Viviana
;; REGISTRATION NUMBER: 30,930
;; REFERENCE/DOCKET NUMBER: P66 38215
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 213-622-7700
;; TELEFAX: 213-489-4210
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1384 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
US-08-162-402B-1
Alignment Scores:
Pred. No.: 1,73e-24 Length: 1384
Score: 346.00 Matches: 84
Percent Similarity: 47.97% Conservative: 34
Best Local Similarity: 34.15% Mismatches: 94
Query Match: 9.97% Indels: 35
DB: 2 Gaps: 5
US-10-060-830-1114 (1-653) x US-08-162-402B-1 (1-1384)
Qy 158 LeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMetGlu 177
Db 145 CTGGCTTTGAGTACTGGGCTGTGAGCTGAACGGATCGCGCAATCCCTGGGCTGAAG 204
Qy 178 SerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGlu---TrpThrAsp 196
Db 205 AATAACAGCATCCCTGACAGCAGATCAGCGCTCCAGCAGCTACAAGACCTGGGGCTTG 264
Qy 197 HisThrGlyGlnGluAsnSerTrpLysProLysAlaArgLeuLysLysProGly--- 215
Db 265 CATCTC-----TTACGTGAACCCCTCTATCAGCGGTGGACAGCAGGGGCAAC 315
Qy 216 ---ProProTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGlnIleAspLeuAsn 234
Db 316 TTCAACGGCTGGTGGGGAGGTACGGTAACATCAGTGGCTGCAGTGGACCTGGGC 375
Qy 235 LysGluLysLysIleThrGlyIleThrGlySerThrMetValGluHisAsnTyr 254
Db 376 TCCTCGAAGGAGTGACAGGCATCATCACCGAGGGGCGCGTAACCTTTGGCTCTCTCCAG 435
Qy 255 TyrValSerAlaTyrArgIleLeuTyrSerAspAspGlyGlnLysTrpThrValTyrArg 274
Db 436 TTTGTGGCATCTACAAGTTGGCTACAGTAATGACAGTGGGAAGTGGAGTACGATACCAG 495

Qy 275 GluProGlyValIleGlnAspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspVal 294
Db 496 GACCCAGCAGCTGGCAGCAGTAAGATCTTCCTGGCAACTGGGCAACCACTCCCAAG 555
Qy 295 ArgAsnAsnPheLeuProProIleIleAlaArgPheIleArgValAsnProThrGlnTyr 314
Db 556 AAGAAGCTTTTGTAGAGCGCCATCTGGCTGCGTATGTGGCATCTGCTGTAGCCTGG 615
Qy 315 GlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPheIlePro--- 331
Db 616 CACAACCGCATCGCCTGGCGCTGGAGCTCTGGGCTGTAGTGGCCACCTGCCACCCC 675
Qy 331 ----- 331
Db 676 AGGTCTTCTGCTTTCCATGGCGCGCTGCTCTTGGCTTCTCAGCCCTTTAAATCAC 735
Qy 332 -----LysGlyArgProProLysLeuThrGlnProProProProArgAsnSer 347
Db 736 ATAGGCTGGGAGCTGGGGAAGGGAGGG-TGTTTCAGAGGCGCAGCACCACACAGTCAC 794
Qy 348 AsnAspLeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAlaProLysPhe 367
Db 795 CCCTCCCTCCCTCTTCCACACCTCCACCTCTCACGGGCGCTGCCCGAGCCCTAAG--- 851
Qy 368 ThrGlnProLeuGlnPro 373
Db 852 CCCGTCCTCCCTAAACCC 869
RESULT 12
US-08-162-402B-7
; Sequence 7, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1934 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

Qy 251 GluHisAsnTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTrp 270
 Db 4171 CTTACCAAGTGTAGTAAAGAGTCTCATCTCCAGCAGTCAAGATGGCCATCAGTGG 4230
 Qy 271 ThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyr 290
 Db 4231 ACTCTCTTTTTCAGAAATGGC-----AAAGTAAAGGTTTTTCAGGGAAATCAAGACTCC 4284
 Qy 291 HisGlnAspValArgAsnAsnPheLeuProIleIleAlaAArgPheIleArgValAsn 310
 Db 4285 TTCACACTGGTGAACCTCTAGACCCACCGTTACTGACTCGTACCTTGAATTCAC 4344
 Qy 311 ProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPhe---- 329
 Db 4345 CCCAGAGTTGGTGCACCAAGATTGCCCTGAGGATGGAGTTCTGGCTCGCAGGCACAG 4404
 Qy 330 -----IleProLysGlyArgProProLys 337
 Db 4405 GACCTCTACTGAGGTGGCCACTGCAGCAGCTGCCACTGCCCTCCTCCTCCTCAG 4464
 Qy 338 LeuThrGlnProProPro 344
 Db 4465 CTCAGGGCAGTGTCCCTCCC 4485

RESULT 14

US-08-717-294-41

; Sequence 41, Application US/08717294

; Patent No. 6114148

; GENERAL INFORMATION:

; APPLICANT: SEED, BRIAN

; APPLICANT: HAAS, JURGEN

; TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF

; PROTEINS

; NUMBER OF SEQUENCES: 110

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Clark & Elbing LLP

; STREET: 176 Federal Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/717,294

; FILING DATE: 20-SEP-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Elbing, Karen L

; REGISTRATION NUMBER: 35,238

; REFERENCE/DOCKET NUMBER: 00786/345001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-428-0200

; TELEFAX: 617-428-7045

; TELEX:

; INFORMATION FOR SEQ ID NO: 41:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4670 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-717-294-41

Alignment Scores:

Pred. No.:

Score:

2.84e-21

322.00

Length:

Matches:

4670

72

Percent Similarity: 50.72% Conservative: 33
 Best Local Similarity: 34.78% Mismatches: 80
 Query Match: 9.28% Indels: 22
 DB: 3 Gaps: 4

US-10-060-830-1114 (1-653) x US-08-717-294-41 (1-4670)

Qy 153 SerValValGlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGly 172
 Db 3912 AGCATTCGACGACCTCTTCGCATGTCAGTGGCTGTGATTTAAATAGTTGCAGCATG 3971
 Qy 173 ThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerSerValLeu 192
 Db 3972 CCATTTGGGAATGGAGTAAGCAATATCAGATGCACAGATTACTGCTTCATCCCTACTTT 4031
 Qy 193 GluTyrThrAspHisThrGlyGlnGluAsnSerTyrLysProLysLysAlaArgLeuLys 212
 Db 4032 -----ACCAATATGTTTGCACCTGCTCTCTTCAAAAGCTGACTTCAC 4076
 Qy 213 LysProGly-----ProProTrpAlaAlaPheAlaThrAspGluTyrGlnTrpLeuGln 230
 Db 4077 CTCCAAGGGAGGAGTAATGCCTGGAGCCTCAGGTGAATAATCAAAAGAGTGGCTGCAA 4136
 Qy 231 IleAspLeuAsnLysGluLysIleThrGlyIleIleThrThrGlySerThrMetVal 250
 Db 4137 GTGGACTTCCAGAGACAATGAAGTCACAGGAGTAACCTACTCAGGGAGTAAATCTCTG 4196
 Qy 251 GluHisAsnTyrTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTrp 270
 Db 4197 CTTACCAAGTGTAGTGAAGGAGTTCCTCATCTCAGCAGTCAAGATGGCCATCAGTGG 4256
 Qy 271 ThrValTyrArgGluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyr 290
 Db 4257 ACTCTCTTTTTCAGAAATGGC-----AAAGTAAAGTTCAGGAAATCAAGACTCC 4310
 Qy 291 HisGlnAspValArgAsnAsnPheLeuProIleIleAlaAArgPheIleArgValAsn 310
 Db 4311 TTCACACCTGTGTGAACCTCTTAGACCCACCGTTACTGACTCGTACCTTGAATTCAC 4370
 Qy 311 ProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuGlyCysGlnPhe--- 329
 Db 4371 CCCAGAGTTGGTGCACCAAGATTGCCCTGAGGATGCTGGGCTGCGAGGCACAG 4430
 Qy 330 -----IleProLysGlyArgProProLys 337
 Db 4431 GACCTCTACTGAGGTGGCCACTGCAGCAGCTGCCACTGCCCTCCTCCTCCTCAG 4490
 Qy 338 LeuThrGlnProProPro 344
 Db 4491 CTCAGGGCAGTGTCCCTCCC 4511

RESULT 15

US-08-366-851A-1

; Sequence 1, Application US/08366851A

; Patent No. 5681746

; GENERAL INFORMATION:

; APPLICANT: Bodner, Mordechai

; APPLICANT: De Polo, Nicolas J.

; APPLICANT: Hsu, David Chi-Rang

; APPLICANT: Chang, Steven

; TITLE OF INVENTION: Retroviral Delivery of Full Length Factor VIII

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Viagene, Inc.

; STREET: 11055 Roselle Street

; CITY: San Diego

; STATE: California

; COUNTRY: U.S.A.

; ZIP: 92121

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/366,851A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Chambers, Daniel M.
REGISTRATION NUMBER: 34,561
REFERENCE/DOCKET NUMBER: 930049.438
TELEPHONE: (619) 452-1288
TELEFAX: (619) 452-2616
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8967 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 110..7165
US-08-366-851A-1

Alignment Scores:
Pred. No.: 8.11e-21 Length: 8967
Score: 322.00 Matches: 72
Percent Similarity: 50.72% Conservatives: 33
Best Local Similarity: 34.78% Mismatches: 80
Query Match: 9.28% Indels: 22
DB: 1 Gaps: 4

US-10-060-830-1114 (1-653) x US-08-366-851A-1 (1-8967)

Qy 153 SerValValGlyHisLeuSerThrSerLeuPheThrPhelySerGlyCysTyrGly 172
Db 6635 AGCATTCGCAGCACTCTTCGCATGGAGTTGATGGCTGATTAATAAGTAGTCAGCATG 6694
Qy 173 ThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerValLeu 192
Db 6695 CCATTGGGATGGAGATGAAGCAATATACATGACAGATTACTGCTTCTACTCTTT 6754
Qy 193 GluTrpThrAspHisThrGlyGlnGluAsnSerTrpLysProLysLysAlaArgLeuLys 212
Db 6755 -----ACCAATATCTTCCACCTGGTCTCTCCAAAAGCTCGACTTCAC 6799
Qy 213 LysProGly-----ProProTrpAlaAlaPheAlaThrAspGluTrpGlnTrpLeuGln 230
Db 6800 CTCAAGGAGGAGTAATGCTGGAGACCTCAGGTGAATAATCCAAAAGAGTGGCTGCAA 6859
Qy 231 IleAspLeuAsnLysGluLysIleThrGlyIleIleThrThrGlySerThrMetVal 250
Db 6860 GTGACTTCCAGAGACAAATGAAGTACAGAGAGTAACTACTCAGGAGTAAATCTCTG 6919
Qy 251 GluHisAsnTrpValSerAlaTrpArgIleLeuTrpSerAspGlyGlnLysTrp 270
Db 6920 CTTACAGCANGATGTGAAGAGTTCTCTCATCTCCAGCATCAAGATGCCATCAGTGG 6979
Qy 271 ThrValTrpArgGluProGlyValGluAspLysIlePheGlnGlyAsnLysAspTyr 290
Db 6980 ACTCTCTTTTTCAGATGGC-----AAAGTAAAGGTTTTTCAGGGAATCAAGACTCC 7033
Qy 291 HisGlnAspValArgAsnAsnPheLeuProPheIleAlaArgPheIleArgValAsn 310
Db 7034 TTCACACTGTGGTGAATCTCTAGACCCCGTTACTGACTCGCTACCTTCGAATTCAC 7093
Qy 311 ProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGlnPhe--- 329
Db 7094 CCCCAGAGTTGGTGCACCATGATTCCTGAGGATGGAGTTCTGGGCTCGGAGGCACAG 7153
Qy 330 -----IleProLysGlyArgProLys 337
Db 7154 GACCTTACTGAGGTGGCCACTGCAGCACCTGCGCTACCTCCCTCTCCCTCCTCAG 7213

Qy 338 LeuThrGlnProProPro 344
Db 7214 CTCAGGGCAGTGTCCCTCCC 7234
Search completed: January 21, 2003, 12:56:40
Job time : 80.5905 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:51:36 ; Search time 7.56006 Seconds
(without alignments)
1717.007 Million cell updates/sec

Title: US-10-060-830-1114
Perfect score: 3470
Sequence: 1 MFLFLLLVLLVLLLEDAGA.....TQEVSGAGRGDCDVFKEIL 653

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap:*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pap:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	873.5	25.2	715	9 US-10-003-132-2	Sequence 2, Appli
2	417	12.0	503	9 US-10-003-132-4	Sequence 4, Appli
3	416	12.0	503	10 US-09-823-038A-51	Sequence 51, Appl
4	347	10.0	2224	9 US-10-115-563-14	Sequence 14, Appl
5	324	9.3	458	9 US-10-003-132-6	Sequence 6, Appli
6	317	9.1	1431	12 US-10-095-718-4	Sequence 4, Appli
7	308.5	8.9	1438	12 US-10-006-091-1	Sequence 1, Appli
8	308.5	8.9	1438	12 US-10-047-257-1	Sequence 1, Appli
9	308.5	8.9	1471	12 US-10-095-718-2	Sequence 2, Appli
10	308.5	8.9	2332	9 US-09-957-641-2	Sequence 2, Appli
11	295	8.5	931	12 US-10-104-440-4	Sequence 4, Appli
12	265	7.6	923	12 US-10-104-440-2	Sequence 2, Appli
13	220.5	6.4	756	9 US-09-992-598-62	Sequence 62, Appl
14	220.5	6.4	756	9 US-09-989-293A-62	Sequence 62, Appl
15	220.5	6.4	756	9 US-09-989-735-62	Sequence 62, Appl
16	220.5	6.4	756	9 US-09-990-444-62	Sequence 62, Appl
17	220.5	6.4	756	9 US-09-989-730-62	Sequence 62, Appl
18	220.5	6.4	756	9 US-09-990-436-62	Sequence 62, Appl
19	220.5	6.4	756	9 US-09-991-181-62	Sequence 62, Appl

20	220.5	6.4	756	9 US-09-993-687-62	Sequence 62, Appl
21	220.5	6.4	756	9 US-09-989-734-62	Sequence 62, Appl
22	220.5	6.4	756	9 US-10-028-072-392	Sequence 392, App
23	220.5	6.4	756	9 US-09-997-653-62	Sequence 62, Appl
24	220.5	6.4	756	10 US-09-989-722-62	Sequence 62, Appl
25	220.5	6.4	756	10 US-09-989-723-62	Sequence 62, Appl
26	220.5	6.4	756	10 US-09-989-279-62	Sequence 62, Appl
27	220.5	6.4	756	10 US-09-989-727-62	Sequence 62, Appl
28	220.5	6.4	756	10 US-09-989-731-62	Sequence 62, Appl
29	220.5	6.4	756	10 US-09-989-732-62	Sequence 62, Appl
30	220.5	6.4	756	10 US-09-991-073-62	Sequence 62, Appl
31	220.5	6.4	756	10 US-09-990-442-62	Sequence 62, Appl
32	220.5	6.4	756	10 US-09-991-163-62	Sequence 62, Appl
33	220.5	6.4	756	10 US-09-993-604-62	Sequence 62, Appl
34	220.5	6.4	756	10 US-09-990-456-62	Sequence 62, Appl
35	220.5	6.4	756	10 US-09-989-721-62	Sequence 62, Appl
36	220.5	6.4	764	10 US-09-900-518A-2	Sequence 2, Appli
37	200.5	5.8	855	10 US-09-771-161A-196	Sequence 196, App
38	198	5.7	734	9 US-10-174-590-458	Sequence 458, App
39	198	5.7	734	9 US-10-176-758-458	Sequence 458, App
40	198	5.7	734	12 US-10-052-586-458	Sequence 458, App
41	185	5.3	913	10 US-09-223-490-4	Sequence 4, Appli
42	183	5.3	520	10 US-09-771-161A-105	Sequence 105, App
43	182.5	5.3	250	10 US-09-770-643A-8	Sequence 8, Appli
44	182.5	5.3	279	10 US-09-770-643A-10	Sequence 10, Appl
45	182.5	5.3	534	10 US-09-770-643A-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-10-003-132-2
; Sequence 2, Application US/10003132
; Publication No: US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-132-2

Query Match	25.2%;	Score 873.5;	DB 9;	Length 715;
Best Local Similarity	35.2%;	Pred. No. 3.8e-63;		
Matches 225;	Conservative 102;	Mismatches 209;	Indels 103;	Gaps 22;
Qy	13	ILLEDAGAQQKCYGGLGLOMNHIESKGNITLLEMSGTHVSGRGLASVYIDKQDLIT	72	
Db	99	LLTSSDDQYPCYG-SMTVPKELLNTSEVTVRFESGSHISGRGFLTYASSDHPDLIT	157	
Qy	73	CLDTASNFLEPEFSKYPACGLLPFAEISOTIPHGVRDPSPLCMAGVHAGVSVNTLGGQI	132	
Db	158	CLERASHYLKTEYSKFCPCAGCRVAGDISGNMVDGYEDTSLCKAAIHAGIIDEAGQI	217	
Qy	133	SVVISKGIPYESSLANNTSVVGHLSLFTFTKTSCTGLMGESGVITADPOITASSVL	192	
Db	218	SVLQKRGISRYEGILANGVLSDGLSDKRLFTNSGCRSLSE----	273	
Qy	193	EWTDTGQNSKPKARLKKPPWAAATFDE----	248	
Db	274	QSVNESGDQVHWSPGQARLQDQGPSWASGDSNNHKKPREMLEIDLGEKKKITITGTST	333	

Pred. No.: 5.63e-104	Length: 1272
Score: 1151.50	Matches: 218
Percent Similarity: 91.16%	Conservative: 9
Best Local Similarity: 87.55%	Mismatches: 21
Query Match: 33.18%	Indels: 1
DB: 11	Gaps: 1

US-10-060-830-1114 (1-653) x AK006805 (1-1272)

Qy	405	VAlAlaLeuAlaValLeuValProValLeuValMetValLeuThrLeuIleLeu	424
Db	488	GTCCACACTGGCTGCGTTCGTGCTGCATGCCCTCACCACACTCATCTC	547
Qy	425	IleLeuValCysAlaTrpHisTrpArgAsnArgLysLysThrGlUGlyThrTyrrasp	444
Db	548	ATTCTAGTGTGTGCTGCACTGGAGAACAGGAAAGAAACTGAAGGCGCCATGAT	607
Qy	445	LeuProTyrrTPASrArgAlaGlyTrpTrpLysGlyMetLysGlnPheLeuProAlalys	464
Db	608	TTACCCCACCTGGATCGGCGAGTGTGGTGAAGAAGAAATGAAGCTTCTCCCTCCCAAG	667
Qy	465	AlaValAspHisGluGluThrProValArgTyrrSerSerSerGluValAsnHisLeuSer	484
Db	668	TCGCTGGACACGAGGAGACCCGAGTGGCTACAGCACTAGTAGTGAGTCACTCACCTGAGT	727
Qy	485	ProArgGluValThrThrValLeuGlnAlaAspSerAlaGluTyrrAlaGlnProLeuVal	504
Db	728	GCCAGGGAAGTCACACACAGTCTGCAGCCGACCTCTGCAGAATAATGCACAGCCCCCTCGTG	787
Qy	505	GlyGlyIleValGlyThrLeuHisGlnArgSerThrPheLysProGluGluGlyLysGlu	524
Db	788	GGAGGAATTGTGGCACACTCCATCAGAGATCCACCTTTAAACCTTCAGGAAGGGAAGAA	847
Qy	525	AlaGlyTyrrAlaAspLeuAspProTyrrAsnSerProGlyGlnGluValTyrrHisAlaTyrr	544
Db	848	CGAGCTATGCAGACCTCGATCTTACAACTCTCCAATGCAGGAAGTGTACCCGCCCTAT	907
Qy	545	AlaGluProLeuProIleThrGlyProGluTyrrAlaThrProIleMetAspMetSer	564
Db	908	GCTCAACCACTTGCCTTAACCGGGCTGAGTACGCAACCCGATCGTCATGGACATGCA	967
Qy	565	GlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsn	584
Db	968	GGGACCCCAACAGCTTGGTGGTGTGCTGCCCTCCACATCCACCTTCAAACCTCAGGGAGC	1027
Qy	585	GlnProProProLeuValGlyThrTyrrAsnThrLeuLeuSerArgThrAspSercysSer	604
Db	1028	CAGCTCAGCTTAGTGGGAACTTACAACTCTTCTCTCCAGCACTGACAGCTGTCTCC	1087
Qy	605	SerAlaGlnAlaGlnTyrrAspThrProLysAlaGlyLysProGlyLeuProAlaProAsp	624
Db	1088	TCAGCCAGGCTCAGTATGACACCCCAAAAAGGTGGGAAG---TCAGCTGCTACCCACAG	1144
Qy	625	GluLeuValTyrrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgaspGly	644
Db	1145	GAATGGTATACAGGTCGCCAGAGCCAGGAGCTATCAGGAGCAGGAAGGATGAG	1204
Qy	645	GluCysAspValPheLysGluIleLeu	653
Db	1205	AAGTTTGATGCTTTTAAGAATCCCTT	1231

RESULT 5
A1435602/c
LOCUS tH79c10.xl Soares_NHHMPu_S1 Homo sapiens cDNA clone IMAGE:2124882
DEFINITION 3' similar to tr:Q14089 Q14089 HYPOTHETICAL PROTEIN t.; mRNA sequence.
ACCESSION A1435602
VERSION A1435602.1 GI:4304334
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (Bases 1 to 641)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 865 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 450
Location/Qualifiers
1..641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2124882"
/clone_lib="Soares_NHHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, an,
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7m3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uteru
NbHPU, and fetal heart NBHH19W) were mixed, and ss cl
were made in vitro. Following HAP purification, this D
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pool
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 161 a 125 c 153 g 202 t
ORIGIN

Alignment Scores:
Pred. No.: 2.56e-98 Length: 641
Score: 1090.50 Matches: 209
Percent Similarity: 98.60% Conservative: 2
Best Local Similarity: 97.66% Mismatches: 3
Query Match: 31.43% Indels: 1
DB: 9 Gaps: 0

US-10-060-830-1114 (1-653) x A1435602 (1-641)

Qy	219	AlaAlaPheAlaThrAspGluTyrrGlnTrpLeuGlnIleaspLeuAsnLysGluLysLys	238
Db	641	CTCGCTTTTCCCTCTCAATACCAGTGGTTACAAATAGATTGTGAATAGGAAAAA	582
Qy	239	IleThrGlyIleThrThrGlySerThrMetValCluHisAsnTyrrValSerAla	258
Db	581	ATAACAGGCAATATATACCCACTGGAT--ACCATGGTGAGCACAACTTACTATGTCTGCC	524
Qy	259	TyrArgIleLeuTyrrSerAspGlyGlnLysTrpThrValTyrrArgGluProGlyVal	278
Db	523	TACAGAAATCTGTACAGTATGATGGCAGAAATGGAGTGTGTACAGAGAGCCGTGTGTG	464
Qy	279	GluGlnAspLysIlePheGlnGlyAsnLysAspTyrrHisGlnAspValArgAsnAsphe	298
Db	463	GAGCAAGATAAGATATTTCAAGGAAACAAAGATTATCACCAGGATGTGCGTAATACTTT	404
Qy	299	LeuProProIleIleAlaArgPheIleArgValAsnProThrGlnTrpGlnLysIle	318
Db	403	TGCGCCACCAATATTCCACGTTTTATTAGAGTGAATCTTACCCCAATGGCAGCAGAAAAT	344
Qy	319	AlaMetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeu	338
Db	343	GCCATGAAATGGAGCTGCTCGGATGTCAGTTTATCTTAAAGGTGCTCTCTCCAAAACTT	284
Qy	339	ThrGlnProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLys	358
Db	283	ACTCAACCTCCACCTCTCGGAACAGCAATAGCTTCAAAACACTACAGCCCTCCAAAA	224

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM9756 row: e column: 18
 High quality sequence stop: 618.
 Location/Qualifiers

FEATURES

source
 1..618
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3922169"
 /clone_lib="NIH_MGC_72"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Skin; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
 Average insert size 2 kb. Library constructed by Life
 Technologies."
 BASE COUNT 201 a 135 c 139 g 143 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.06e-95 Length: 618
 Score: 1064.00 Matches: 202
 Percent Similarity: 98.06% Conservative: 0
 Best Local Similarity: 98.06% Mismatches: 3
 Query Match: 30.66% Indels: 2
 DB: 12 Gaps: 0

US-10-060-830-1114 (1-653) x BR894226 (1-618)

Qy 161 SerLeuphethrPheLysThrSerGlyCysTyrglyThrLeuGlyMetGluSerGlyVal 180
 |||||
 Db 3 AGCTTTTACATTAAACAGAGTGGATGTTATGCAACACTGGGATGGGCTGGTG 62
 Qy 181 IleAlaaspProGlnIleThrAlaSerValLeuGluTrpThrAspHisThrGlyGln 200
 |||||
 Db 63 ATCCGGATCTCAATAACAGCATCTGCTGCTGGAGTGGACTGACCACAGGGCAA 122
 Qy 201 GluAsnSerTrpLysProLysLysAlaArgLeuLysLysProGlyProTrpAlaAla 220
 |||||
 Db 123 GAGAACAGTGGAAACCCAAAGCCAGGCTGAAACACCTGGACCCCTGGTGCT 182
 Qy 221 PheAlaThrAspGluTyrglnTrpLeuGlnIleAspLeuAsnLysGluLysLysIleThr 240
 |||||
 Db 183 TTTGCCACTGATGATACCACTGGTTACAAATAGATTGTAATAAGGAAAGAAATAACA 242
 Qy 241 GlyIleIleThrThrGlySerThrMetValGluHisAsnTyrtVrValSerAlaTyArg 260
 |||||
 Db 243 GGCATTATACCACTGGATCCACCATGGTGGAGCACAAATTAATATATGTCTGCCTACAGA 302
 Qy 261 IleLeutySerAspGlyGlnLysTrpThrValTyArgGluProGlyValGluGln 280
 |||||
 Db 303 ATCTGTACAGTGATGGGCAGAAATGGACTGTGTACAGAGAGCCCTGGTGTGGAGCAA 362
 Qy 281 AspLysIlePheGlnGlyAsnLysAspTyrtHisGlnAspValArgAsnAsnPheLeuPro 300
 |||||
 Db 363 GATAAGATATTTCAGGAAACAAAGATTATCACCAGGATGTGCGTAATAACTTTTGGCA 422
 Qy 301 -ProfileIleAlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIleAlaMe 320
 |||||
 Db 423 CCCAATTATTCACGCTTTTATAGAGTGATCTTACCCCAATGGCAGCAGAAATGCCAT 482
 Qy 320 tLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProProLysLeuThrGln 340
 |||||
 Db 483 GAAATGAGCTGCTCGGATGTCAGTTTATCTCCFAAAGGTGCTCTCCCAACATCTACTCA 542
 Qy 340 nProProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLysIleAl 360
 |||||
 Db 543 ACT-CCACCTCTCGGAACAGCAATGACCTCAAAAACACTACAGCCCTCCAAAAATAGC 601

Qy 360 aLysGlyArgAlaPro 365
 |||||
 Db 602 CAAAGGCTGCTGCCCC 617
 RESULT 8
 AAL47037/c
 LOCUS
 DEFINITION
 2032a05.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
 IMAGE:588560 3' similar to TR:G704441 G704441 HYPOTHETICAL 40.0 KD
 PROTEIN,, mRNA sequence.
 ACCESSION AAL47037
 VERSION AAL47037.1 GI:1716444
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 671)
 AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
 Chissole,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
 M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
 B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
 Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisan,E.,
 Underwood,K., Wohlmann,P., Waterston,K., Wilson,R. and Marra.M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 COMMENT
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewartson.wustl.edu
 This clone is available royalty-free through LLNL: contact the
 IMAGE Consortium (infoimage.llnl.gov) for further information.
 Seq primer: -40M13 fwd. from Amersham
 High quality sequence stop: 439.
 FEATURES
 source
 1..671
 /organism="Homo sapiens"
 /db_xref="GDB:4620862"
 /db_xref="taxon:9606"
 /clone="IMAGE:588560"
 /clone_lib="Stratagene colon (#937204)"
 /tissue_type="tumor"
 /cell_line="T84 carcinoma cell line"
 /lab_host="SOLR cells (kanamycin resistant)"
 /note="Organ: colon; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dr. T-84 colonic epithelial cell line. Average
 insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
 sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTT 3".
 BASE COUNT 176 a 162 c 133 g 198 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5.29e-94 Length: 671
 Score: 1047.50 Matches: 210
 Percent Similarity: 95.89% Conservative: 0
 Best Local Similarity: 95.89% Mismatches: 8
 Query Match: 30.19% Indels: 4
 DB: 9 Gaps: 1
 US-10-060-830-1114 (1-653) x AAL47037 (1-671)
 Qy 97 PheAlaGluIleSerGlyThrIleProHisGlyTyArgAspSerProLeuCysMet. 116
 |||||
 Db 651 TTTTCTGAGATATCTGGACC-ATTCCCTCATGATATGA-GATTCTCGCCA---TGTCNT 597
 Qy 117 AlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlnIleSerValIle 136
 |||||
 Db 596 GGCTGGTCTCATGCAGGAGTAGTGTCAAAACACGTT-GGCGGCCCAATCAGTGTGTGAATT 538

QY 369 AGTGTCACACAGCTGGGGCCCAATCAGTGTGTAATTAAGTATTCCTCTATTA 428
Db 571 CATCAGATGAAGTAGGTGGCCACATCACTGCTTCAGAGCAAGGATAAGTCACTA 630
QY 429 TGAAGTCTCTTTGGTCAACAGCTACATCTGCTGGGACACTTATCTACAGTCTTTT 488
Db 631 TGAAGGACTCTCTGGCAATGGGCTGCTCTCCGGCATGGTCTTTTGTCTGGAAACGGAAT 690
QY 489 TACATTTAAGACAAGTGGATGTTATGGAACACTGGGGATGGAGTCTGGTGTGATC 543
Db 691 TCTTTTACACCCAGGAATGAATATACAACTGTGGGATTCATCAGTGATC 745

RESULT 12
US-09-974-298-167
; Sequence 167, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 167
; LENGTH: 4599
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 235636.1
US-09-974-298-167

Query Match 4.6%; Score 89.8; DB 9; Length 4599;
Best Local Similarity 50.7%; Pred. No. 2e-16;
Matches 242; Conservative 0; Mismatches 232; Indels 3; Gaps 1;
QY 507 ATGTTATGGAACACTGGGGATGGAGTCTGGTGTGATCGGGGATCCTCAATAACACATC 566
Db 784 ATGCTAGGCCCCACTGGGAATGAAGTGAATATATCAACACGAGCAATACAGCTTC 843
QY 567 ATCTGTCTGGAGTGGACTACACACAGGCGCAAGAGACAGTTGGAAACCCCAAAAGC 636
Db 844 CTCTACTCT---ACCGAGCTCTTTTGGACTCCCAAAATGGTATCCCTACTATCCAGTCT 900
QY 627 CAGGCTGAAAAAACCTGGACCGCTTGGGCTGCTTTTGGCAGTGAATACAGTGGTT 686
Db 901 TAATAAGAGGGCTTATAATGCGTGGACAGCTGCAGAAATGACAGATGGCGGTGAT 960
QY 687 ACAATAGATTGAATAAGGAAAGAAATPAACAGGCAATTAACCACTGGATCCACCAT 746
Db 961 TCAGATAAATTTGCAAGGAAATAGAGTGTACTGGTGTGATTACCCCAAGGACCAAG 1020
QY 747 GGTGGACCAATTAATCTATGCTGTCTACAGAACTCTGTACAGTATGATGGGAGAA 806
Db 1021 GATTGGAGCCAGAGTATATAAATCCTACAAAATGGCTACAGTAATGATGGAAGAC 1080
QY 807 ATGACTGTGTACAGAGGCTGTGGTGGAGCAAGATPAAGATATTTCAAGGAAACAAAGA 866
Db 1081 TTGGCAATGTACAAAGTGAAGGCCCAATGAAGACATGGTGTTCGTGGAAACATTGA 1140
QY 867 TTATCAGCAGATGTGGGTATTAATCTTTTGGCCACCAATTAATGTCAGCTTTTATTAGAT 926
Db 1141 TAACAACACTCCATATGCTAACTCTTTTACACACCCCCCAATTAAGCTCAGTATGTAAGACT 1200
QY 927 GAATCCTACCAATGGGAGCAGAAATTTGCCATGAAATGAGCTGCTGGATGCA 983
Db 1201 CTATCCCCAAGTTTGTGCAAGACATATGCACTTTCCGAATGGAATCTTCTGGCTGTGA 1257

RESULT 13
US-09-880-107-2275
; Sequence 2275, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Uwe
; TITLE OF INVENTION: Gene Logic, Inc.
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2275
; LENGTH: 6909
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M16967
US-09-880-107-2275

Query Match 4.4%; Score 86.6; DB 10; Length 6909;
Best Local Similarity 51.7%; Pred. No. 2.4e-15;
Matches 259; Conservative 0; Mismatches 224; Indels 18; Gaps 2;
QY 493 TTTAAGACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATGCGGGATCT 552
Db 6271 TGTAGGTTAAATGGATGTTCCACACCCCTGGGTATGAAAAATGAAAGATAGAAAAAAG 6330
QY 553 CAATAACAGCATCATCTGTGCTGGAGTGGACTGACACACAGGCGCAAGAGACAGTTGG 612
Db 6331 CAATACAGCTCTTTC-----GTTAAGAAATCTTGGTGGGAGATTTACTGG 6378
QY 613 AAACCCAAAAAGAGCGAGCTGAAAAAACCTGGAC-----CGCCTTGGGCTGCTTTTGGC 666
Db 6379 GAACCTTCCGTGCCGCTGTAATGCCAGGAGCTGTGATGCTGGCAAGCCAGGCA 6438
QY 667 ACTGATCAATACACAGTGGTGTACAAATAGATTTGAATAAGGAAAGAAATAACAGGCAAT 726
Db 6439 AACACAATAAGCAGTGGCTAGAAATGTATCTACTCAAGATCAAGAAGATAACGGCAAT 6498
QY 727 ATACCACTGGATCCACCATGGTGGAGCAATTAATCTATGCTGCTGCCTACAGATCCTG 786
Db 6499 ATACACAGGCTGCAAGTCTCTCTCTGAAATGTATGTAAGAGCTATACCATCCAC 6558
QY 787 TACAGTATGATGGCAGAAATGGACTGTGTACAGAGAGCTGTGTGGAGCAAGATAAG 846
Db 6559 TACAGTACAGGAGTGGATGGAATGGAACCAATACAGGCTGAATCCTCCATGGTGACAAG 6618
QY 847 ATATTTCAAGGAAACAAAGATTAATACACAGGATGTGGTGTATTAATCTTTTGGCCACCAAT 906
Db 6619 ATTTTGAAGGAAATTAATAACCAAGGACATGTGAAGAACTTTTTCACACCCCAATC 6678
QY 907 ATTGCAGTGTATAGAGTGAATCCTTACCAATGGCAGCAGAAATTTGCCATGAAATG 966
Db 6679 ATTTCCAGGTTTATCCGTGTCTCTCTCTGAAATGTATGTAAGAGCTATACCATCCAC 6738
QY 967 GAGCTGCTCGGATGTCAAGTT 987
Db 6739 GAATCTTTGGCTGTGATTT 6759

RESULT 14
US-10-115-563-13
; Sequence 13, Application US/10115563
; Publication No. US20030008307A1

```
Db 199 PGM----- 201
Qy 336 PKLTQPPPPRNSDLKNTAPPKIAKRAPKFTQPLQPRSSNEFFPAQTEQTASPDIRNT 395
Db 202 ----- 201
Qy 396 TVTPNWKDVALAAYLPVLMVLTLLILVCAMHWRKKKTEGYDLPY----WDRA 451
Db 202 ----NIT-TVAIPSVI--FIALLTGMGIFAIC-----RKRKKGN---PVSADAOKT 245
Qy 452 GWWKMKQFLPAKAVDHEETP--VRYSSSE--VNHLSPREVT---TVLOADSAEYQAQPLVG 505
Db 246 GCKWQIKY-----PPARHQSTFTSYDNEREMTKLDLITSDMADYQOPLMI 293
Qy 506 GIVGLTHORSTPKP-----EE-----GKRGAYADLPYNSPGQEVYHAYAEPLPTGPEYA 556
Db 294 GRGTVARKGSTPRMDTDEVRVNTASGH-----YDCPHRPRGRHEVALPLTHSEPEYA 348
Qy 557 TPIL-----MDMSCHPTTSYGQBSTSTFKATGNQPPPLVCTYNTLLSRTDSCSA 606
Db 349 TPIVERHLLRAHTFTSQGYRVPG-PRPTHKHSXGGFFPATGATQVESYQRPASPKPV 407
Qy 607 QAOYDTPKAGKPLPAPDELVYQVQSTQEVSGAGRDG 644
Db 408 GGYDKP-AASSFLDSRD-----PASQSQMTSGDDG 438

RESULT 6
US-10-095-718-4
; Sequence 4, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: Canine B-domain deleted factor VIII
US-10-095-718-4

Query Match 9.1%; Score 317; DB 12; Length 1431;
Best Local Similarity 33.1%; Pred. No. 2.6e-17;
Matches 84; Conservative 34; Mismatches 96; Indels 40; Gaps 7;

Qy 100 ISGTPHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKI-----PYESSLANNVY 152
Db 1206 LDGNKWHYSRGNS-----TGLTVFNGVNDSSGIKHNFNPPIIAQYIRLHPT 1253
Qy 153 --SVVGHLSLSTFTKTSCTGLGMSGVIAADPQITASSVLEWTDHTGOENSWKPKAR 210
Db 1254 HYSIRSTLRWELLCDFNSCSPLGMSKAIQITASSYL-----SSMLATWSQSAR 1308
Qy 211 LKKPG--PPWAAAFATDEYQWLQIDLNKEKKITGITGTSTWVEHNYYSAYRILYSDGQ 268
Db 1309 LHLOGRTNARVQANNPKEWLQVDFRKTMTKVTGTTQGVKSLISMYKFEFLISSSQDGH 1368
Qy 269 KTVVYREPGEVDKIFQGNKDHQDVRRNPLPPIIARFIRVNPQWQOQKIAMKMLLGCO 328
```

```
Db 1369 NWTFLQNG--KVKVFGNRRDSSFTVRNLEPPLVARYVRLHPQSWAHIALRLEVLGCD 1426
Qy 329 FIPKGRPPKLTQPP 342
Db 1427 -----TQQP 1430

RESULT 7
US-10-006-091-1
; Sequence 1, Application US/10006091
; Patent No. US20020102730A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.1
; CURRENT APPLICATION NUMBER: US/10/006,091
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
US-10-006-091-1

Query Match 8.9%; Score 308.5; DB 12; Length 1438;
Best Local Similarity 37.1%; Pred. No. 1.3e-16;
Matches 66; Conservative 31; Mismatches 72; Indels 9; Gaps 3;

Qy 153 SVVGHLSLSTFTKTSCTGLGMSGVIAADPQITASSVLEWTDHTGOENSWKPKARLK 212
Db 1263 SIRSTLRWELMCCDLNCSNPLGMSKAIQITASSYF-----TMMFATWSFKARLH 1317
Qy 213 KPG--PPWAAAFATDEYQWLQIDLNKEKKITGITGTSTWVEHNYYSAYRILYSDGQKW 270
Db 1318 LQGRSNARVQANNPKEWLQVDFRKTMTKVTGTTQGVKSLISMYKFEFLISSSQDGHQW 1377
Qy 271 TVVYREPGEVDKIFQGNKDHQDVRRNPLPPIIARFIRVNPQWQOQKIAMKMLLGCO 328
Db 1378 TLEFQNG--KVKVFGNQRDSSFTPVVNSLDPLLTLYRLIRHPQSWVHQAIALRMEVLGCE 1433

RESULT 8
US-10-047-257-1
; Sequence 1, Application US/10047257
; Patent No. US20020115152A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.2
; CURRENT APPLICATION NUMBER: US/10/047,257
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
US-10-047-257-1
```

```
Query Match      8.9%; Score 308.5; DB 12; Length 1438;
Best Local Similarity 37.1%; Pred. No. 1.3e-16;
Matches 66; Conservative 31; Mismatches 72; Indels 9; Gaps 3;

QY 153 SVVGHLSLFTFKTSCGYGLGMSGVIAADPOITASSVLEWTDHTGOENSKPKKARLK 212
DB 1263 SIRSTLRMELMGCDLNSCMLPGMESKASDAQITASSYF-----TNMFATWSPSKARLH 1317
QY 213 KPG--PPWAAAFATDEYQWLQIDLNKEKKITGIITGTSTMVEHNYVVSAYRILYSDDGQKW 270
DB 1318 LQGRSNARWPOVNNPKEMQLQVDFQKTMKVGTGTQGVKSLLTSMYVKEFLISSQDGHOW 1377
QY 271 TVYREPGVEQDKIFQGNKYHODVNRNLPPIIARFIRVNPQTQOQKIAMKMWELGCG 328
DB 1378 TLFQNG--KVKVFGNQDSFTPVVNSLDPLLRILYRIHPQSWVHQIALRMEVLGCE 1433

RESULT 9
US-10-095-718-2
; Sequence 2, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; FILE REFERENCE: Methods of Using the Same
; CURRENT APPLICATION NUMBER: US/10/095,718
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/159,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1471
; TYPE: PRT
; ORGANISM: Homo sapiens B-domain deleted factor VIII
; FEATURE:
; OTHER INFORMATION: Homo sapiens BDD FVIII
US-10-095-718-2

Query Match      8.9%; Score 308.5; DB 12; Length 1471;
Best Local Similarity 37.1%; Pred. No. 1.3e-16;
Matches 66; Conservative 31; Mismatches 72; Indels 9; Gaps 3;

QY 153 SVVGHLSLFTFKTSCGYGLGMSGVIAADPOITASSVLEWTDHTGOENSKPKKARLK 212
DB 1296 SIRSTLRMELMGCDLNSCMLPGMESKASDAQITASSYF-----TNMFATWSPSKARLH 1350
QY 213 KPG--PPWAAAFATDEYQWLQIDLNKEKKITGIITGTSTMVEHNYVVSAYRILYSDDGQKW 270
DB 1351 LQGRSNARWPOVNNPKEMQLQVDFQKTMKVGTGTQGVKSLLTSMYVKEFLISSQDGHOW 1410
QY 271 TVYREPGVEQDKIFQGNKYHODVNRNLPPIIARFIRVNPQTQOQKIAMKMWELGCG 328
DB 1411 TLFQNG--KVKVFGNQDSFTPVVNSLDPLLRILYRIHPQSWVHQIALRMEVLGCE 1466

RESULT 10
US-09-957-641-2
; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
```

```
; CURRENT FILING DATE: 2001-09-16
; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-957-641-2

Query Match      8.9%; Score 308.5; DB 9; Length 2332;
Best Local Similarity 37.1%; Pred. No. 2.6e-16;
Matches 66; Conservative 31; Mismatches 72; Indels 9; Gaps 3;

QY 153 SVVGHLSLFTFKTSCGYGLGMSGVIAADPOITASSVLEWTDHTGOENSKPKKARLK 212
DB 2157 SIRSTLRMELMGCDLNSCMLPGMESKASDAQITASSYF-----TNMFATWSPSKARLH 2211
QY 213 KPG--PPWAAAFATDEYQWLQIDLNKEKKITGIITGTSTMVEHNYVVSAYRILYSDDGQKW 270
DB 2212 LQGRSNARWPOVNNPKEMQLQVDFQKTMKVGTGTQGVKSLLTSMYVKEFLISSQDGHOW 2271
QY 271 TVYREPGVEQDKIFQGNKYHODVNRNLPPIIARFIRVNPQTQOQKIAMKMWELGCG 328
DB 2272 TLFQNG--KVKVFGNQDSFTPVVNSLDPLLRILYRIHPQSWVHQIALRMEVLGCE 2327

RESULT 11
US-10-104-440-4
; Sequence 4, Application US/10104440
; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Quan
; TITLE OF INVENTION: ANTAGONISTS OF NEUROPILIN RECEPTOR FUNCTION AND USE
; FILE REFERENCE: 48802 C
; CURRENT APPLICATION NUMBER: US/10/104,440
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/580,803
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/069,155
; PRIOR FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 60/069,687
; PRIOR FILING DATE: 1997-12-29
; PRIOR APPLICATION NUMBER: 60/078,541
; PRIOR FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 931
; TYPE: PRT
; ORGANISM: human
US-10-104-440-4

Query Match      8.5%; Score 295; DB 12; Length 931;
Best Local Similarity 23.0%; Pred. No. 8.6e-16;
Matches 133; Conservative 72; Mismatches 246; Indels 128; Gaps 21;

QY 23 GKVCGLGLQMN--HSIESKGNEITLLFMSGIHVSRGFLASVSVIDKQDLITCLDTASN 79
DB 102 GKHC-----NIAPPTIISSGSMLYIKFTSDYARQAGFSLRKEIFKTS-----EDCSKN 152
QY 80 FLEPEFSKYCPAGCLLPFAEISGTIPHYRDSRSPCLCMAGVHAGVSVNTLGGQIS-----133
DB 153 FTSPNGTIESPG-----FPE---KYPHNL-DCTFTILAKPKMEIILQFLIFLDEHPLQV 203
QY 134 -----VWISKGIP-----YYESSLANNVTSVVGHLSLFT-----164
```

```

Db 204 GEGDCKYDMLDMDGIPHYGPIGKYCGKTPSELRSSTGLSLTFPHDMAYAKGFSAR 263
Oy 155 -----FKTSGCYGTGMEGCVIADPOITASSVLEMTDHTGQENSKPKKARLKKRCP 216
Db 264 XYLHQBEPLENFQCNVPLGMEGRIANEQISASS-----TSDGRWTPQOOSRLHGDN 316
Oy 217 PAAATATDEYOMLQIDINKKKTITGTTG--STMVHNYYSARILYSDGOKMTYVR 274
Db 317 GTPMIDSKKEYLQVDFLMTLTAIOGAISRETONGYKYSKLEYSTIGEDMMYR 376
Oy 275 EGVGEODKIFQGNKDYHODVRNNFLPRLIARFIRVNPOTMOOKIMKMKELGCOFIPKRG 334
Db 377 HG--KNHKVFQANNDATVYLNKLHAPLITFRVRIAPQWHSGLRLLEFGC-----427
Oy 335 PPKLTQPPPPRNSNDLKTTPAPKIAKGRAPKF-----TOPLOPPSSNEFPAPQETQTAS 389
Db 428 --RYTDAFPCSNMGLSGLIADSOISASSTOEYLSWSPSARLVSSSGFPRIPQAQPE 485
Oy 390 PDIRMTVTPNVTKDY-----ALAAVLVPLVMVLTLLILVCAWHM-----RNR 435
Db 486 EWLQVLDLGTPTKVTYITIGARGGDSITAVEARAFYRKVKYSYSLNGKWEYIODEPTQO 545
Oy 436 KKTTEGT--YDLRYMDRAGMMKMOFLPAKAVDHEETPVRYSSSEVNHLSPREVTVLQ 493
Db 546 PRLFEGNMHYDP-----DIRRDP-----IPAQYRVYPERKWSAGIGMRLE 588
Oy 494 ADSAEY--AOLYGVIGVTLHORSTFKP---EEGKEAG 526
Db 589 VLGDWTDGSKPYVETLGPVKSEETTPYPTFEATEGCG 627

RESULT 12
US-10-104-440-2
; Sequence 2, Application US/10104440
; Patent No. US20020132774A1
; GENERAL INFORMATION:
; APPLICANT: KLAGSBRUN, Michael
; APPLICANT: SOKER, Shay
; APPLICANT: MIAO, Hua Quan
; TITLE OF INVENTION: ANTAGONISTS OF NEUROFILIN RECEPTOR FUNCTION AND USE
; FILE REFERENCE: 48802 C
; CURRENT APPLICATION NUMBER: US/10/104,440
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/580,803
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/069,155
; PRIOR FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: 60/069,687
; PRIOR FILING DATE: 1997-12-29
; PRIOR APPLICATION NUMBER: 60/078,541
; PRIOR FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 923
; TYPE: PRT
; ORGANISM: human
US-10-104-440-2

Query Match 7.6%; Score 265; DB 12; Length 923;
Best Local Similarity 26.3%; Pred. No. 2,4e-13;
Matches 102; Conservative 50; Mismatches 138; Indels 98; Gaps 16;

Oy 17 DACAGGKCYGGLGLOMNHSTESKNEITLLFMGSHVSGGLASYSVIDK-----67
Db 219 DVCPHGRIC--GQKTPGRSSSGILSNVFTYDLSIAIEGFSANYSVLQSSSEDFKCM 276
Oy 68 -----ODLITCLDPTASNFLPEFSKYCPACGLLPFAEISGTIP--HGVRDSSPL 114
Db 277 EALGMESGEIHSQITASSQSYTNMSAERSR-----LNYPE-NCWGTGEDSYREMIGY 328
Oy 115 CMAGVHAGVSNLTGGOISVYISKGIPTYESSLANNVTS-----153

```

```

Db 329 DLGLRFTVANGTGAISKEKTK--KYYVKYKIDVSSNGEDMTITEKGNKPVLPQNGT 385
Oy 154 -----VGHLSLSTLT-----FKTSG-----CYTGLGMEGCVIADP 184
Db 386 NPTDVAVVAFPRPLTFRVRIKIPATWENGISMRFEYGGKITDYPCSGMLGWSGLSDS 445
Oy 185 QITASSVLEMTDHTGQENSKPKKARLKKRPPNA-APATDEY--QWQIDINKKKTITG 241
Db 446 QITSS-----NOGDRN--WMPENIRLVTSRSGMALPPAPHSYINEMQIDIGEEKIYNG 497
Oy 242 ITTGSTVHEHNYYSARYILYSDGOKWTYRREGVEODKIFQGNKDYHODVRNNFLPP 301
Db 498 IITGSKIRENVYPAKRFYIGYSNNKSDMKMLMDSKRAKSFEGNNNTDPELNTF-PA 556
Oy 302 IIAFIRVNPQ--WQOKIAMKMLGCG 328
Db 557 LSTRFIRIYPERATHGGLRMLLGE 584

RESULT 13
US-09-992-598-62
; Sequence 62, Application US/09992598
; Patent No. US20020160384A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Bolstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC20
; CURRENT APPLICATION NUMBER: US/09/992,598
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106

```

1	PRIOR APPLICATION NUMBER: 60/089801
2	PRIOR FILING DATE: 1998-06-18
3	PRIOR APPLICATION NUMBER: 60/089907
4	PRIOR FILING DATE: 1998-06-18
5	PRIOR APPLICATION NUMBER: 60/089908
6	PRIOR FILING DATE: 1998-06-18
7	PRIOR APPLICATION NUMBER: 60/089947
8	PRIOR FILING DATE: 1998-06-19
9	PRIOR APPLICATION NUMBER: 60/089948
10	PRIOR FILING DATE: 1998-06-19
11	PRIOR APPLICATION NUMBER: 60/089952
12	PRIOR FILING DATE: 1998-06-19
13	PRIOR APPLICATION NUMBER: 60/090246
14	PRIOR FILING DATE: 1998-06-22
15	PRIOR APPLICATION NUMBER: 60/090252
16	PRIOR FILING DATE: 1998-06-22
17	PRIOR APPLICATION NUMBER: 60/090254
18	PRIOR FILING DATE: 1998-06-22
19	PRIOR APPLICATION NUMBER: 60/090340
20	PRIOR FILING DATE: 1998-06-23
21	PRIOR APPLICATION NUMBER: 60/090355
22	PRIOR FILING DATE: 1998-06-23
23	PRIOR APPLICATION NUMBER: 60/090429
24	PRIOR FILING DATE: 1998-06-24
25	PRIOR APPLICATION NUMBER: 60/090431
26	PRIOR FILING DATE: 1998-06-24
27	PRIOR APPLICATION NUMBER: 60/090435
28	PRIOR FILING DATE: 1998-06-24
29	PRIOR APPLICATION NUMBER: 60/090444
30	PRIOR FILING DATE: 1998-06-24
31	PRIOR APPLICATION NUMBER: 60/090445
32	PRIOR FILING DATE: 1998-06-24
33	PRIOR APPLICATION NUMBER: 60/090472
34	PRIOR FILING DATE: 1998-06-24
35	PRIOR APPLICATION NUMBER: 60/090535
36	PRIOR FILING DATE: 1998-06-24
37	PRIOR APPLICATION NUMBER: 60/090540
38	PRIOR FILING DATE: 1998-06-24
39	PRIOR APPLICATION NUMBER: 60/090542
40	PRIOR FILING DATE: 1998-06-24
41	PRIOR APPLICATION NUMBER: 60/090557
42	PRIOR FILING DATE: 1998-06-24
43	PRIOR APPLICATION NUMBER: 60/090676
44	PRIOR FILING DATE: 1998-06-25
45	PRIOR APPLICATION NUMBER: 60/090678
46	PRIOR FILING DATE: 1998-06-25
47	PRIOR APPLICATION NUMBER: 60/090690
48	PRIOR FILING DATE: 1998-06-25
49	PRIOR APPLICATION NUMBER: 60/090694
50	PRIOR FILING DATE: 1998-06-25
51	PRIOR APPLICATION NUMBER: 60/090695
52	PRIOR FILING DATE: 1998-06-25
53	PRIOR APPLICATION NUMBER: 60/090696
54	PRIOR FILING DATE: 1998-06-25
55	PRIOR APPLICATION NUMBER: 60/090862
56	PRIOR FILING DATE: 1998-06-26
57	PRIOR APPLICATION NUMBER: 60/090863
58	PRIOR FILING DATE: 1998-06-26
59	PRIOR APPLICATION NUMBER: 60/091360
60	PRIOR FILING DATE: 1998-07-01
61	PRIOR APPLICATION NUMBER: 60/091478
62	PRIOR FILING DATE: 1998-07-02
63	PRIOR APPLICATION NUMBER: 60/091544
64	PRIOR FILING DATE: 1998-07-01
65	PRIOR APPLICATION NUMBER: 60/091519
66	PRIOR FILING DATE: 1998-07-02
67	PRIOR APPLICATION NUMBER: 60/091626
68	PRIOR FILING DATE: 1998-07-02
69	PRIOR APPLICATION NUMBER: 60/091633
70	PRIOR FILING DATE: 1998-07-02
71	PRIOR APPLICATION NUMBER: 60/091978
72	PRIOR FILING DATE: 1998-07-07
73	PRIOR APPLICATION NUMBER: 60/091982
74	PRIOR FILING DATE: 1998-07-07

```

; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match      6.4%; Score 220.5; DB 9; Length 756;
Best Local Similarity 31.0%; Pred. No. 7.5e-10;
Matches 58; Conservative 33; Mismatches 69; Indels 27; Gaps 7;

QY 174 LGMESGVADPQITASSVLEW--TDHTGQENSWKPKARLKKP---GPPWAFAFATDEYQW 228
||:| | | : ||:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 139 LGLETLKITDFOLHASTVKRYGLGAHRGLNI---QAQINENDFYDCAWCAGRNDLQQW 194
||:| | | : ||:| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 229 LQIDLNKEKIIGTITGTSTWMEHYNYSAIRLYSDDGQKW-TVYRPFGVEQDKIFQGN 287
:::| | | : ||:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 IEVDARRLTRFTGVTIQGRNSLWLSDWVTSYKMVSNSDSHTWTVTYKNGSG---DMIEFEGN 251

QY 288 KDYHQDVNRNFLPLPIARTIRVNPTQW--QQKIANKMELLGCQFIKPRGPLLTQPPPPR 345
| | | | | | : ||:| | | : ||:| | | | | | | | | | | | | | | | | | |
Db 252 SEKEIPVLNELPVPMVARYIRINPOSFEDNGSICMRMELGCL-----PDEN 299
| | | | | | : ||:| | | : ||:| | | | | | | | | | | | | | | | | | |

QY 346 NSMDLN 352
| | | | | | : ||:| | | : ||:| | | | | | | | | | | | | | | | | | |
Db 300 NYHRRN 306
| | | | | | : ||:| | | : ||:| | | | | | | | | | | | | | | | | | |

RESULT 14
US-09-989-293A-62
; Sequence 62, Application US/09989293A
; Patent No. US20020177164A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC66
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20

```

```

; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02

```

```

; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match      6.4%; Score 220.5; DB 9; Length 756;
Best Local Similarity 31.0%; Pred. No. 7.5e-10;
Matches 58; Conservative 33; Mismatches 69; Indels 27; Gaps 7;

QY 174 LCMESGVADPOITASSVLEW--TDHTGOENSMKPKARKLP---GPPWAFATDEYQM 228
DB 139 LLEFLKRTDFOLHASTVRYGLGAHGRHLN1-----QACINENDRYDGMCAGRNDLQW 194
QY 229 LOIDLNKEKRTIGITTSFWENHYSAVRLYSDDQKM--TYRREPGVQDRIPOGN 287
DB 195 IEVDARRLTRFTGVLTQGNLSLMSDWTSYKVMNSDSHTWTVKNGSG---DMIFEEN 251
QY 288 KYHODVRRNNEFLPPIARIRVNPOTW--QOKIAKMKELGCOFLPKGRPKLTQPPPPR 345
DB 252 SEKEIPVLNELPVPVWVARYRINPOSWFNDGSIChRMETLGCPL-----PDEN 299
QY 346 NSNDLKN 352
DB 300 NYTHRRN 306

RESULT 15
US-09-989-735-62
; Sequence 62, Application US/09989735
; Publication No. US20020193299A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730PIC61
; CURRENT APPLICATION NUMBER: US/09/989,735
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770

```

3	PRIOR APPLICATION NUMBER: 60/089538
4	PRIOR FILING DATE: 1998-06-17
5	PRIOR APPLICATION NUMBER: 60/089598
6	PRIOR FILING DATE: 1998-06-17
7	PRIOR APPLICATION NUMBER: 60/089599
8	PRIOR FILING DATE: 1998-06-17
9	PRIOR APPLICATION NUMBER: 60/089600
10	PRIOR FILING DATE: 1998-06-17
11	PRIOR APPLICATION NUMBER: 60/089653
12	PRIOR FILING DATE: 1998-06-17
13	PRIOR APPLICATION NUMBER: 60/089801
14	PRIOR FILING DATE: 1998-06-18
15	PRIOR APPLICATION NUMBER: 60/089907
16	PRIOR FILING DATE: 1998-06-18
17	PRIOR APPLICATION NUMBER: 60/089908
18	PRIOR FILING DATE: 1998-06-18
19	PRIOR APPLICATION NUMBER: 60/089947
20	PRIOR FILING DATE: 1998-06-19
21	PRIOR APPLICATION NUMBER: 60/089948
22	PRIOR FILING DATE: 1998-06-19
23	PRIOR APPLICATION NUMBER: 60/089952
24	PRIOR FILING DATE: 1998-06-19
25	PRIOR APPLICATION NUMBER: 60/090246
26	PRIOR FILING DATE: 1998-06-22
27	PRIOR APPLICATION NUMBER: 60/090252
28	PRIOR FILING DATE: 1998-06-22
29	PRIOR APPLICATION NUMBER: 60/090254
30	PRIOR FILING DATE: 1998-06-22
31	PRIOR APPLICATION NUMBER: 60/090349
32	PRIOR FILING DATE: 1998-06-23
33	PRIOR APPLICATION NUMBER: 60/090355
34	PRIOR FILING DATE: 1998-06-23
35	PRIOR APPLICATION NUMBER: 60/090429
36	PRIOR FILING DATE: 1998-06-24
37	PRIOR APPLICATION NUMBER: 60/090431
38	PRIOR FILING DATE: 1998-06-24
39	PRIOR APPLICATION NUMBER: 60/090435
40	PRIOR FILING DATE: 1998-06-24
41	PRIOR APPLICATION NUMBER: 60/090444
42	PRIOR FILING DATE: 1998-06-24
43	PRIOR APPLICATION NUMBER: 60/090445
44	PRIOR FILING DATE: 1998-06-24
45	PRIOR APPLICATION NUMBER: 60/090472
46	PRIOR FILING DATE: 1998-06-24
47	PRIOR APPLICATION NUMBER: 60/090542
48	PRIOR FILING DATE: 1998-06-24
49	PRIOR APPLICATION NUMBER: 60/090535
50	PRIOR FILING DATE: 1998-06-24
51	PRIOR APPLICATION NUMBER: 60/090540
52	PRIOR FILING DATE: 1998-06-24
53	PRIOR APPLICATION NUMBER: 60/090542
54	PRIOR FILING DATE: 1998-06-24
55	PRIOR APPLICATION NUMBER: 60/090557
56	PRIOR FILING DATE: 1998-06-24
57	PRIOR APPLICATION NUMBER: 60/090676
58	PRIOR FILING DATE: 1998-06-25
59	PRIOR APPLICATION NUMBER: 60/090678
60	PRIOR FILING DATE: 1998-06-25
61	PRIOR APPLICATION NUMBER: 60/090690
62	PRIOR FILING DATE: 1998-06-25
63	PRIOR APPLICATION NUMBER: 60/090694
64	PRIOR FILING DATE: 1998-06-25
65	PRIOR APPLICATION NUMBER: 60/090695
66	PRIOR FILING DATE: 1998-06-25
67	PRIOR APPLICATION NUMBER: 60/090696
68	PRIOR FILING DATE: 1998-06-25
69	PRIOR APPLICATION NUMBER: 60/090862
70	PRIOR FILING DATE: 1998-06-26
71	PRIOR APPLICATION NUMBER: 60/090863
72	PRIOR FILING DATE: 1998-06-26
73	PRIOR APPLICATION NUMBER: 60/091360
74	PRIOR FILING DATE: 1998-07-01
75	PRIOR APPLICATION NUMBER: 60/091478
76	PRIOR FILING DATE: 1998-07-02
77	PRIOR APPLICATION NUMBER: 60/091544

```
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match      6.4%; Score 220.5; DB 9; Length 756;
Best Local Similarity 31.0%; Pred. No. 7.5e-10;
Matches 58; Conservative 33; Mismatches 69; Indels 27; Gaps 7;

QY 174 LGHESGVIADPQITASSVLEW--TDHTGQENSWKPKKARKLP---GPPWAAAFATDEYQW 228
Db 139 LGLETLKITDFOQLHASTVKRYGLGAHRGLNI----QAGINENDFYDGAWCAGRNDLQOW 194

QY 229 LOJDLNKEKITGIITGSTMVEHNYVVSAYRILYSDDGOKW-TVYREPGVEODKIFOGN 287
Db 195 IEVDARLRTRFTGVITQGRNSLMSDWVTSIKVWVSNDSTWTVTKNGSG---DMIFEGN 251

QY 288 KQYHQDVRRNFPPIIARFIRVNPTQW--OOKIAMKMEILGCOFIPKGRPPKLTQPPPPR 345
Db 252 SEKEIPVLNLPVPMVARYIRINPQSWFDNGSICMRMEILGCPL-----PDPN 299

QY 346 NSNDLKN 352
Db 300 NYVHRN 306
```

Search completed: January 21, 2003, 09:55:09
Job time : 13.5601 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	1897.4	96.7	5657	9	US-09-974-298-96	Sequence 96, Appl
2	544.6	27.8	640	10	US-09-879-536-480	Sequence 480, Appl
3	473.8	24.1	580	10	US-09-864-761-12400	Sequence 12400, A
4	467	23.8	467	10	US-09-864-761-30344	Sequence 30344, A
5	252.8	12.9	2145	9	US-10-003-132-12	Sequence 12, Appl
6	204	10.4	3151	9	US-10-003-132-11	Sequence 11, Appl
7	141.8	7.2	1509	9	US-10-003-132-13	Sequence 13, Appl
8	119	6.1	2836	9	US-10-003-132-3	Sequence 3, Appl
9	117.4	6.0	1871	10	US-09-823-038A-43	Sequence 43, Appl
10	115.4	5.9	1374	9	US-10-003-132-14	Sequence 14, Appl
11	92.6	4.7	2868	9	US-10-003-132-5	Sequence 5, Appl
12	89.8	4.6	4599	9	US-09-974-298-167	Sequence 167, Appl
13	86.6	4.4	6909	10	US-09-880-107-2275	Sequence 2275, Ap
14	86.6	4.4	6925	9	US-10-115-563-13	Sequence 13, Appl
15	86.6	4.4	6925	9	US-10-115-563-26	Sequence 26, Appl
16	86.6	3.9	1270	10	US-09-880-107-3020	Sequence 3020, Ap
17	64.6	3.3	970	10	US-09-215-450-14	Sequence 14, Appl
18	63.4	3.2	4629	9	US-09-150-811-7	GENERAL INFORMAT
19	63.4	3.2	4999	9	US-10-007-968-14	Sequence 14, Appl

	Query Match	96.7%	Score 1897.4	DB 9	Length 5657
	Best Local Similarity	99.7%	Pred. No. 0		
	Matches 1901	Conservative 0	Mismatches 6	Indels 0	Gaps 0
Qy	56	GAGCCACGAGGCAATAC	TGCGTCTGGGGTTGCAATCAACATTC	CAATTGAATCAA	115
Db	284	GAATGAAATAGGCAATAC	TGCGTCTGGGGTTGCAATGAACATTC	CAATTGAATCAA	343
Qy	116	AAGCAATGAATCACATTC	TGCTTCATGAGTGGAAATCCATGTTTCTCGACGCGGATTTT		175
Db	344	AAGCAATGAATCACATTC	TGCTTCATGAGTGGAAATCCATGTTTCTCGACGCGGATTTT		403
Qy	176	TGGCCTCATCTCTGTTAT	AGATAAACAAGATCTAATTACTTGTGTTGGACACTGCATCCA		235
Db	404	TGGCCTCATCTCTGTTAT	AGATAAACAAGATCTAATTACTTGTGTTGGACACTGCATCCA		463

236 ATTTTGGAACTGAGTGTAGTACTGCCCCAGCTGTTGTCGCTTCTTTTGGCTG 295
|||||
Db 464 ATTTTGGAACTGAGTGTAGTACTGCCCCAGCTGTTGTCGCTTCTTTTGGCTG 523
|||||
Oy 296 AGATATCGAACAATTCCTCATGATATAGAGATTCCTCGCAATGTGCATGGCTGTG 355
|||||
Db 524 AGATATCGAACAATTCCTCATGATATAGAGATTCCTCGCAATGTGCATGGCTGTG 583
|||||
Oy 356 TGCATGACAGAGTGTGTCAACACGTTGGCGGCCAAATAGTGTCTTAATAGTAAG 415
|||||
Db 584 TGCATGACAGAGTGTGTCAACACGTTGGCGGCCAAATAGTGTCTTAATAGTAAG 643
|||||
Oy 416 GTATTCCTTATTAAGAGTCTTTGGCTAACAACTGTCATCTGTGGGACACTTAT 475
|||||
Db 644 GTATTCCTTATTAAGAGTCTTTGGCTAACAACTGTCATCTGTGGGACACTTAT 703
|||||
Oy 476 CTACAAGTCTTTTACATTTAAGACAACTGATGTTATGAACTGTGGGATGAGTCTG 535
|||||
Db 704 CTACAAGTCTTTTACATTTAAGACAACTGATGTTATGAACTGTGGGATGAGTCTG 763
|||||
Oy 536 GTGTGATCGGAGTCTCTCAATTAACAGCATCTCTGTGTGTGGAGTGCACACAG 595
|||||
Db 764 GTGTGATCGGAGTCTCTCAATTAACAGCATCTCTGTGTGTGGAGTGCACACAG 823
|||||
Oy 596 GGCAGAGAACAGTTGGAAACCCAAAGGCGAGGCTGAAAAACCTGAGCCCTTGGG 655
|||||
Db 824 GGCAGAGAACAGTTGGAAACCCAAAGGCGAGGCTGAAAAACCTGAGCCCTTGGG 883
|||||
Oy 656 CTGCTTTTCCACTGATGATACAGTGTGTTACAAATAGATTTGAATAGAAAAAGAAA 715
|||||
Db 884 CTGCTTTTCCACTGATGATACAGTGTGTTACAAATAGATTTGAATAGAAAAAGAAA 943
|||||
Oy 716 TAAACGCAATTAATTAACCACTGATCCACATGTTGGAGCAATTAATTAATGTCGCT 775
|||||
Db 944 TAAACGCAATTAATTAACCACTGATCCACATGTTGGAGCAATTAATTAATGTCGCT 1003
|||||
Oy 776 ACAGATCCGTGACGTGATGATGAGCAGAAATGACCTGTACAGAGAGCCCTGTGTG 835
|||||
Db 1004 ACAGATCCGTGACGTGATGATGAGCAGAAATGACCTGTGTACAGAGAGCCCTGTGTG 1063
|||||
Oy 836 AGCAGATAGATATTTCAAGAAACAAAGATTAATCAACAGATGTGCGTAATACTTT 895
|||||
Db 1064 AGCAGATAGATATTTCAAGAAACAAAGATTAATCAACAGATGTGCGTAATACTTT 1123
|||||
Oy 896 TGGCACCATTATTTGACGTTTATTAGATGATGATCTCAATGCGACAGAAAAATGG 955
|||||
Db 1124 TGGCACCATTATTTGACGTTTATTAGATGATGATCTCAATGCGACAGAAAAATGG 1183
|||||
Oy 956 CCATGAATAAGAGCTGCTGATGATGATTTATTCCTAAAGGTCTCTCAAAACTTA 1015
|||||
Db 1184 CCATGAATAAGAGCTGCTGATGATGATTTATTCCTAAAGGTCTCTCAAAACTTA 1243
|||||
Oy 1016 CTCAACTCCACTCTCTGAGAACAGCATATACCTCAAAAAACATACAGCCCTCCAAAA 1075
|||||
Db 1244 CTCAACTCCACTCTCTGAGAACAGCATATACCTCAAAAAACATACAGCCCTCCAAAA 1303
|||||
Oy 1076 TAGCCAAAGTGTGCCCCAAATTTAGCAACACTACAACTCGAGAGTACGAATGAT 1135
|||||
Db 1304 TAGCCAAAGTGTGCCCCAAATTTAGCAACACTACAACTCGAGAGTACGAATGAT 1363
|||||
Oy 1136 TTCTGACAGACAGAACAAACAACTGCAATCTGATATCAGAAATACGCTAACTC 1195
|||||
Db 1364 TTCTGACAGACAGAACAAACAACTGCAATCTGATATCAGAAATACGCTAACTC 1423
|||||
Oy 1196 CAATATGTAACCAAGATGATGAGCTGCTGATGTTCTTGTCTGTGTGATGATGCTC 1255
|||||
Db 1424 CAATATGTAACCAAGATGATGAGCTGCTGATGTTCTTGTCTGTGTGATGATGCTC 1483
|||||
Oy 1256 TCACATCTCTATCTCTATTAATGATGATGATGATGATGATGATGATGATGATGAT 1315
|||||
Db 1484 TCACATCTCTATCTCTATTAATGATGATGATGATGATGATGATGATGATGATGAT 1543
|||||
Oy 1316 CTGAAGGACCACTATGATCTTACTGTGAGCCGAGGCTTGTGTAAGGATGAAAGC 1375

|||||
Db 1544 CTGAAGGACCACTATGATCTTACTTACTGTGAGCCGAGGCTGTGTAAGAAATGAAAGC 1603
|||||
Oy 1376 AGTTCTCTCTGCAAAACAGAGTGAACCAATGAGAAACCCCACTGCTATACAGACAGC 1435
|||||
Db 1604 AGTTCTCTCTGCAAAACAGAGTGAACCAATGAGAAACCCCACTGCTATACAGACAGC 1663
|||||
Oy 1436 AAGTTAATCAGCTGAGTCCAGAGAAATGATCAGACAGTGTGTGAGGCTGATCTGAGAGT 1495
|||||
Db 1664 AAGTTAATCAGCTGAGTCCAGAGAAATGATGATCAGACAGTGTGTGAGGCTGATCTGAGAGT 1723
|||||
Oy 1496 ATGCTCAGCACTGTGTGAGAGAAATGTTGTGATCACTTCAATCAAGATATACCTTTAAAC 1555
|||||
Db 1724 ATGCTCAGCACTGTGTGAGAGAAATGTTGTGATCACTTCAATCAAGATATACCTTTAAAC 1783
|||||
Oy 1556 CAGAGAAAGAAAGAGAGAGGCTATGACAGCTAGATTCCTTACAACTACAGAGGACAG 1615
|||||
Db 1784 CAGAGAAAGAAAGAGAGGCTATGACAGCTAGATTCCTTACAACTACAGAGGACAG 1843
|||||
Oy 1616 AAGTTATCATGCTATGCTGAACCACTCCCAATTAACGGGCTGTGATATGCAACCCCAA 1675
|||||
Db 1844 AAGTTATCATGCTATGCTGAACCACTCCCAATTAACGGGCTGTGATATGCAACCCCAA 1903
|||||
Oy 1676 TCATCATGACATGTACAGGCAACCCCAACTTCAATGCTGAGCCCTCCACATCCACTT 1735
|||||
Db 1904 TCATCATGACATGTACAGGCAACCCCAACTTCAATGCTGAGCCCTCCACATCCACTT 1963
|||||
Oy 1736 TCAGAGCTACGGGGAACCAACTCCCACTAGTGGGAACCTTACAAATACACTTCTCTCA 1795
|||||
Db 1964 TCAGAGCTACGGGGAACCAACTCCCACTAGTGGGAACCTTACAAATACACTTCTCTCA 2023
|||||
Oy 1796 GGAAGTACAGCTCTCTCAGCCCAAGCCCAAGTATGATATACCCGGAAGCTGGGAAGCAG 1855
|||||
Db 2024 GGAAGTACAGCTCTCTCAGCCCAAGCCCAAGTATGATATACCCGGAAGCTGGGAAGCAG 2083
|||||
Oy 1856 GTTACTGCCCCAGACGAATGTTGTATGACAGGTGACAGACAGACAGCAAGATATCAG 1915
|||||
Db 2084 GTTACTGCCCCAGACGAATGTTGTATGACAGGTGACAGACAGACAGCAAGATATCAG 2143
|||||
Oy 1916 GAGCAGGAAGGAGATGGGGAATGTGATGTTTAAAGAAATCTTTGA 1962
|||||
Db 2144 GAGCAGGAAGGAGATGGGGAATGTGATGTTTAAAGAAATCTTTGA 2190

RESULT 2
US-09-879-536-480/c
; Sequence 480, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilsson O.
; APPLICANT: Steilmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Carino, Theodore J.
; APPLICANT: Dertli, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/879, 536
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/088, 801
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Query Match      27.8%; Score 544.6; DB 10; Length 640;
Best Local Similarity 96.7%; Pred. No. 5.2e-155;
Matches 616; Conservative 0; Mismatches 15; Indels 6; Gaps 6;

QY 893 TTTTGGCCACCAATTATTGCACGCTTTTATTAGAGTGAATCTCTCCCAATGGCAGCAGAAAA 952
Db      |||||||
QY 632 TTTTGGCCACCAATTATGGCAGC-TTTATTAGAGT-AATCCTTACCATGGCAGCAG-AAA 576
Db      |||||||
QY 953 TTGCCATGAATGGAGCTGCTCGGATGTCAGTTTATTCTTAAAGTCTCTCTCCAAAC 1012
Db      |||||||
QY 575 TTGCCATGAAT-GAAGCTGCTCGGATGTCAGTTTATTCTTAAAC-CTNCTCTCCAAAC 518
Db      |||||||
QY 1013 TTACTCAAGCTCCAGCTCTCGGACAGCAATGACCTTCAAAACACTACAGCCCTCCAA 1072
Db      |||||||
QY 517 TTACTCAAGCTCCAGCTCTTNGAAGCAATGCTTCAAAACCACTACAGCCCTCC-A 459
Db      |||||||
QY 1073 AATAGCCAAAGTGTGCCCCCAAAATTTACGCAACCACTACAACTCCGAGTAGCAATG 1132
Db      |||||||
QY 458 AATAGCCAAAGTGTGCCCCCAAAATTTACGCAACCACTACAACTTCGAGTAGCAATG 399
Db      |||||||
QY 1133 AATTTCTCTCAGACAGACAGAAACAACTGCCAGTCTCTGATATCAGAAATACTACCGTAA 1192
Db      |||||||
QY 398 AATTTCTCTCAGACAGACAGAAACAACTGCCAGTCTCTGATATCAGAAATACTACCGTAA 339
Db      |||||||
QY 1193 CTCCAATGTAAACCAAGATGTAGCGCTGGCTGCGAGTTCTTGTCTCCCTGCTGTCATGG 1252
Db      |||||||
QY 338 CTCCAATGTAAACCAAGATGTAGCGCTGGCTGCGAGTTCTTGTCTCCCTGCTGTCATGG 279
Db      |||||||
QY 1253 TCCCTCACTACTCTCATCTCATATTAAGTGTGCTGGCACTGGGACAGAAACAGAA 1312
Db      |||||||
QY 278 TCCCTCACTACTCTCATCTCATATTAAGTGTGCTGGCACTGGGACAGAAACAGAA 219
Db      |||||||
QY 1313 AAACCTGAAGGACCTATGACTTACCTTACTGGACCGGCGAGGTTGGTGGAAAGGAATGA 1372
Db      |||||||
QY 218 AAACCTGAAGGACCTATGACTTACCTTACTGGGCGGCGAGGTTGGTGGAAAGGAATGA 159
Db      |||||||
QY 1373 AGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAAACCCCACTGCGTATAGCAGCA 1432
Db      |||||||
QY 158 AGCAGTTTCTTCTGCAAAAGCAGTGGACCATGAGGAAACCCCACTGCGTATAGCAGCA 99
Db      |||||||
QY 1433 GCGAAGTTTATACCTGAGTCCAGAGAGTACACAGTGTGCGAGGCTGACTCTGCAG 1492
Db      |||||||
QY 98 GCGAAGTTTATACCTGAGTCCAGAGAGTACACAGTGTGCGAGGCTGACTCTGCAG 39
Db      |||||||
QY 1493 AGTATGCTCAGCCACCTGGTAGGAGGAATTTGTTGGTAC 1529
Db      |||||||

RESULT 3
US-09-864-761-12400
; Sequence 12400, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456

```

```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12400
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
US-09-864-761-12400

Query Match      24.1%; Score 473.8; DB 10; Length 580;
Best Local Similarity 99.6%; Pred. No. 1.5e-133;
Matches 475; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1486 TCTGCAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTACACTTCAATCAAGATCT 1545
Db      |||||||
QY 64 TCCACAGAGTATGCTCAGCCACTGGTAGGAGGAATTTGGTACACTTCAATCAAGATCT 123
Db      |||||||
QY 1546 ACCTTTAAACCAAGAAAGAAAGAGCAGGCTATGACAGACCTAGATCTTCAACTCA 1605
Db      |||||||
QY 124 ACCTTTAAACCAAGAAAGAAAGAGCAGGCTATGACAGACCTAGATCTTCAACTCA 183
Db      |||||||
QY 1606 CCAGGGCAGGAAGTTTATCATGCTGCTGTAACCACTCCCAATTCAGGGGCTGAGTAT 1665
Db      |||||||
QY 184 CCAGGGCAGGAAGTTTATCATGCTGCTGTAACCACTCCCAATTCAGGGGCTGAGTAT 243
Db      |||||||
QY 1666 GCACCCCAATCATCATGGACATGTCAGGGCAGCCCACTTTCAGTTGGTCAGCCCTCC 1725
Db      |||||||
QY 244 GCAACCCCAATCATCATGGACATGTCAGGGCAGCCCACTTTCAGTTGGTCAGCCCTCC 303
Db      |||||||
QY 1726 ACATCCCACTTTCAAGGCTACGGGGAACCAACTCCCACTAGTGGGAACCTTCAATACA 1785
Db      |||||||
QY 304 ACATCCCACTTTCAAGGCTACGGGGAACCAACTCCCACTAGTGGGAACCTTCAATACA 363
Db      |||||||
QY 1786 CTTCTCCAGAGTACAGCTGCTCTCAGCCAGCCAGTATGATFACCCCAAGCT 1845
Db      |||||||

```

Wed Jan 22 14:57:21 2003

Db 364 CTTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGCCAGCTATGATATACCCCGAAAGCT 423
QY 1846 GGAAGCCAGGTCTACTCTGCCAGAGGAATGTGTGTTACAGGTGCCACAGACACACAA 1905
Db 424 GGAAGCCAGGTCTACTCTGCCAGAGGAATGTGTGTTACAGGTGCCACAGACACACAA 483
QY 1906 GAAGTATCAGGAGCAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGA 1962
Db 484 GAAGTATCAGGAGCAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTTTGA 540

RESULT 4
US-09-864-761-30344
; Sequence 30344, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30344
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46

; OTHER INFORMATION: EST_HUMAN HIT: AW583777.1, EVALUE 3.00e-51
; OTHER INFORMATION: NT HIT: Z70177.1, EVALUE 2.90e-01
; OTHER INFORMATION: SWISSPROT HIT: Q9Y011, EVALUE 7.40e-01
US-09-864-761-30344

Query Match 23.8% Score 467; DB 10; Length 467;
Best Local Similarity 100.0%; Pred. No. 1.5e-131;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1493 AGTATGCTCAGCCACTGCTAGGAGGAATGTGTGTTACACTTCTATCAAGATCTACCTTTA 1552
Db 1 AGTATGCTCAGCCACTGCTAGGAGGAATGTGTGTTACACTTCTATCAAGATCTACCTTTA 60

QY 1553 AACCAAGAAGGAAAGACAGGCTATGAGACCTTACAGCTTACAACTCACCAGGC 1612
Db 61 AACCAAGAAGGAAAGACAGGCTATGAGACCTTACAGCTTACAACTCACCAGGC 120

QY 1613 AGAAGTTTATCATGCTCTATGTAACCACTCCCAATTTACGGGGCTGAGTATGCAACCC 1672
Db 121 AGAAGTTTATCATGCTCTATGTAACCACTCCCAATTTACGGGGCTGAGTATGCAACCC 180

QY 1673 CAATCATCATGACATGTGAGGGACCCCAACTTCACTTGTGTCAGCCCTCCACATCCA 1732
Db 181 CAATCATCATGACATGTGAGGGACCCCAACTTCACTTGTGTCAGCCCTCCACATCCA 240

QY 1733 CTTTCAAGGCTACGGGGAACCAACTCTCCCACTAGTGGAACTTACAACTACACTTCTCT 1792
Db 241 CTTTCAAGGCTACGGGGAACCAACTCTCCCACTAGTGGAACTTACAACTACACTTCTCT 300

QY 1793 CCAGGACTGACAGCTGCTCTCAGCCAGCCAGCCAGCTATGATACCCGAAAGCTGGGAAGC 1852
Db 301 CCAGGACTGACAGCTGCTCTCAGCCAGCCAGCCAGCTATGATACCCGAAAGCTGGGAAGC 360

QY 1853 CAGGTCTACCTGCCCCCAGACGAATTTGGTGTACAGGTGCCACAGACACACAGAGTAT 1912
Db 361 CAGGTCTACCTGCCCCCAGACGAATTTGGTGTACAGGTGCCACAGACACACAGAGTAT 420

QY 1913 CAGGAGCAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTT 1959
Db 421 CAGGAGCAGGAGGATGGGAATGTGATGTTTTTAAAGAAATCCTT 467

RESULT 5
US-10-003-132-12
; Sequence 12, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG 2CUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2145
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate nucleotide sequence
; NAME/KEY: misc.feature
; LOCATION: (1)...(2145)
; OTHER INFORMATION: n = A,T,C or G
US-10-003-132-12

Query Match 12.9%; Score 252.8; DB 9; Length 2145;
Best Local Similarity 37.0%; Pred. No. 4.1e-66;
Matches 329; Conservative 171; Mismatches 365; Indels 24; Gaps 2;

```
QY 115 AAAGCAATGAATACATCTGCTTATGAGTGAATCCATCTGCTTCTGGACGGGATTT 174
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 370 AAYACWNSGARGTNACNTNGNTYTGARWSNGGWSNCAVATHWSNGNGMGNTTY 429
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 175 TTGGCTCATACTCTGTTATAGATAAACAAGATCTAAATCTACTGTTGGACACTGCATCC 234
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 430 YTNNTNACNTAYGCNWSNWSNGAYCAYCNGGAYTTHACNTGYTNGARMGNGNWSN 489
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 235 AATTTTGGACCTGAGTTAGTACTGCCAGCTGGTGTCTGCTCTCTCTTTGCT 294
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 490 CAYTAYTTHNARCNAGTAYNSNARTYTGCCNGMGNTGYMGNGAYTNGCNGN 549
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 295 GAGATATCTGGACAAATCTCTCATGATATAGATATCTCTGCCAFTGTCATGGCTGT 354
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 550 GAYATHWSNGNAYATGCTNGAYGNTAYMGNGAYACNWSNTYNTNGYARGCNGCN 609
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 355 GTGATGAGGAGTACTGTCAACACAGCTGGGGGGGCAATCAGTGTGTAATAGTAA 414
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 610 ATHCAYCGNGGNATHGCGNAYGARYTNGGNGCARATHWSNNGTNTNCARMGNAAR 669
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 415 GGTATCTCCATATATGAAGTTCTTTGGCTAACACGCTCACATCTGCTGGGACACTTA 474
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 670 GGNATHWSNMGNTAYGARGGNATHYNGCNAAYGGNGTYNWSNMGNGAYGNGWSNYTN 729
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 475 TCTACAAGTCTTTTACATTTAAGACAACTGGATGTATGGAACACTGGGGATGGAGTCT 534
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 730 WNSGAYAAARMGNTYNTTYACNWSNAYGNGTYGWSNMGWSNTYNTWSNTYTGAR--- 786
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 535 GGTGATCGCGGATCCTCAATAACAGCATCACTGTGCTGGAGTGGAGTGCACACACA 594
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 787 -----CCNGAYGNCARATHMGNGCNWSNWSNWSNWSNWSNWSNWSNWSNWSN 837
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 595 GGCAGAGACACAGTTGGAACCCAAAGCCAGCTGAAACAACTGGACCGCTTGG 654
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 838 GNGAYCARGTNCAYTGGWSNCCNGCARGCNMGNTNCARGAYCARGNCCNWSNTGG 897
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 655 GCTGCTTTTCCACTGATGAATACC-----AGTGGTTACAAATAGATTTGAAT 702
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 898 GCNWSNGNGAYWSNWSNAYAYCAVAAACNMGNGARTGGTYNGARATHGAYTNGN 957
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 703 AAGGAAAGAAATACAGGCATTAATAACACTGGATCCACCATGGTGGAGACAAATAC 762
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 958 GARAARAARAATHACNGGNATHMGNACNACNACNACNACNACNACNACNACN 1017
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 763 TATGTCTGCTTACAGATCTCTGACAGTATGGGAGAAATGGACTGTGTACAGA 822
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1018 TATGTNARWSNTYTGATGAATYTAARAAYAAAYWSNARATGGARACNTAYAR 1077
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 823 GAGCCTGGTGGAGCAAGATAAGATATTTCAAGGAAACAAAGATATATCACAGGATGT 882
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1078 GGNATHGNTNAYAYGARGARAARGNTTYCARGGNAAYWSNAYTYTNGMNGAYCCNGTN 1137
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 883 CGTAATAACTTTTGGCACCATAATATGACAGTTTATATAGAGTGAATCTTACCAATGG 942
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1138 CARAAAYAYTYATHCCNCCNATHGNGNMGNTAYGTNGNMGNTNCCNACRACNTGG 1197
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 943 CAGCAGAAAATTCGATGAAAATGGAGCTGCTCGAGTGTCAAGTTTATTC 991
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1198 CAYCARMGNATHGCTNYNARAGTNGARYTNATHGNGTGYCARATHACNC 1246
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
```

RESULT 6

US-10-003-132-1

```
; Sequence 1, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
```

```
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (76)...(2223)
US-10-003-132-1
```

Query Match 10.4%; Score 204; DB 9; Length 3151;

Best Local Similarity 54.1%; Pred. No. 3.4e-51;

Matches 476; Conservative 0; Mismatches 380; Indels 24; Gaps 2;

```
QY 121 AATGAATACATTTGCTTTCATGATGGGAATCCATGTTTCTGCGACCGGATTTTGGCC 180
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 451 AGTGAAGTAACCGCTTGTGAGAGTGGATCCACATTTCTGCGCGGGTTTTGTGCTG 510
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 181 TCATCTCTCTTATAGATAAACAAGATCTAATTAATTGTTTGGACACTGCATCCAAATTT 240
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 511 ACATATGCGAGCAGCGACCATCCAGATTTAATAACATGTTTGAACAGAGTAGCCATTAT 570
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 241 TTGGAACCTGAGTTTCAGTAAGTACTGCCAGCTGGTGTCTGCTTCCTTTTGTGAGATA 300
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 571 TTGAACAGACAATACACAAATTTCTGCCAGCTGGTGTGAGAGAGTAGCAGGAGACATT 630
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 301 TCTGGAACAATTCCTCATGATATAGAGATTCCTCGCCATTTGCGCATGCGTGTGTCAT 360
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 631 TCTGGGAATATGTTAGATGATATAGATACCTCTTTATTGTGCAAGAGTGCATCCAT 690
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 361 GCAGGAGTAGTGTCAACACAGCTGGGCGCCAAATCAATCACTGTGTAATAGTAAAGTATT 420
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 691 GCAGGAGTAATGCTGATGAAGTAGTGGCCAGATAGTGTGCTTACGCCCAAGGATC 750
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 421 CCCTATTATGAAAGTTCTTTGGCTTAACAGCTGCATCTGTGTGGGACACTTATCTACA 480
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 751 AGTCGATATCAAGGATCTCGCAATGGTCTTCTTCGAGGGATGTTCCCTCTCAGAC 810
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 481 AGTCTTTTTCATTTAAGACAAGTGGATGTATGGAACACTGGGGATGGAGTCTGGTGTG 540
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 811 AAGCGATTTCTGTTTACTTCCAATGTTGCAGCAGATCTTGTAGTTTGAACCTTGACGG 870
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 541 ATCCGCGATCCTCAATAACAGCATCATCTGTGTGAGTGGACTGACCAACACAGGCA 600
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 871 -----CAATCAGAGCTTCTTCTCATGCGCAGTGGTCAATGAGAGTGAGAC 918
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 601 GAGAACAGTTTGGAAACCCAAAAGCCAGGCTGAAAACCTGGACCGCTTGGCGTCT 660
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 919 CAAGTTCACCTGGTCTCTCGGCCAAGCCGACTTCAGGACCAAGGCCCATCATGGGCTCG 978
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 661 TTTGCCACTGATGAATACC-----AGTGGTTACAAATAGATTTTCAATAAGAA 708
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 979 GGCACAGATGAGCAACAACACAAACCACGAGAGTGGCTGGAGATCGATTTTGGGGAGAAA 1038
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 709 AAGAAAATAACAGGCATTATTAACCACTGGATCCACCATGGTGGAGCACAATTAATCTGTG 768
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1039 AAGAAAATAACAGGAATTAGGACCCACAGGATCTACACAGTCGAATCTCAACTTTATGTT 1098
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 769 TCTSCCTACAGAACTCTGTACAGTGTATGGGAGAAATGGACTGTGTACAGAGAGCCT 828
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1099 AAGAGTTTGTGTGAATCTTCAAAAACAATAATTTCTAAGTGGAAAGACATATAAAGGAAT 1158
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 829 GGTGTGGAGCAGATAGATATTTCAAGGAACAAGATTTATCACCAGGATGTGGGTAAAT 888
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1159 GTGAATAATGAAGAAAAGGTGTTTCAGGGTAACCTTAACTTTTCGGGACCCAGTGGCAAAAC 1218
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 889 AACTTTTGGCCCAATTTATTGACAGTTTTATTAGAGTGAATCCTTACCAATGGCAGCAG 948
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1219 AATTTCATCCCTCCCATCGTGCCAGATATGTGCGGGTGTGTCGCCACACATGGCACCAG 1278
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
```


; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 1871
; TYPE: DNA
; ORGANISM: Mouse
US-09-823-038A-43

Query Match 6.0%; Score 117.4; DB 10; Length 1871;
Best Local Similarity 54.8%; Pred. No. 4.6e-25;
Matches 232; Conservative 0; Mismatches 191; Indels 0; Gaps 0;
QY 121 AATGAATACATCTGTCATGAGTGAATCCATGTTCTGGACGCGGATTTTGGCC 180
DB 412 AACGAAGTGAAGTCTCTTCAAGAGTGAATCTACATTTCTGGCGGGCTTTCTGCTG 471
QY 181 TCATATCTGTTATAGATAAACAAGATCTAATTAATCTGTTGGACACTGCATCCCAATTTT 240
DB 472 ACCTACGCCAGCAGTGACCATCCAGATTTAATAACCTGTTTGGAAACGAGGACGACATAT 531
QY 241 TTGGAACCTGAGTTCAGTAAGTACTGCCAGCTGTTGCTCTCTCTTCTTCTTCTGAGATA 300
DB 532 TTCTGGAAGAAATATACAGCAAAATTTGCCAGCTGCTGTAGAGACATAGCAGAGATAT 591
QY 301 TCTGGAACAAATTCCTCATGGATATAGAGATTCCTCGCCATTTGCTGCTGCTGTCAT 360
DB 592 TCTGGATACAAAGATGTTACAGAGATACCTCTTATTTGTGCAAACTGCCATCCAC 651
QY 361 GCAGGAGTGTCTCAACACAGTGTGGCGGCCAAATCAGTGTGTTAATTAAGTAAGGTATT 420
DB 652 GCAGGAGTGTCTCAACACAGTGTGGCGGCCAAATCAGTGTGTTAATTAAGTAAGGTATT 420
QY 421 CCTATTATGAAGTCTTTGGCTACACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
DB 712 AGTCACTATGAAGGACTCTCGCCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 771
QY 481 AGTCTTTTACATTTAAGACAAAGTGGATGTTATGGAACACTGGGAGTGGAGTCTGCTGTG 540
DB 772 AGCGATTTCTTTTACACCCAGGAGTGAATATTAACAACTGTGCGGATTCATCAGTG 831
QY 541 ATC 543
DB 832 ATC 834

RESULT 10
US-10-003-132-14
; Sequence 14, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 1374
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate nucleotide sequence
; NAME/KEY: misc_feature
; LOCATION: (1)...(1374)
; OTHER INFORMATION: n = A,T,C or G
US-10-003-132-14

Query Match 5.9%; Score 115.4; DB 9; Length 1374;
Best Local Similarity 34.6%; Pred. No. 1.5e-24;
Matches 140; Conservative 79; Mismatches 186; Indels 0; Gaps 0;
QY 139 TTCATGAGTGGAAATCCATGTTTCTGGACCGGATTTTGGCCCTCATCTCTGTATAGAT 198
DB 232 YTNAAATGAGARWSNAARACNTGYGCNWSNGAYTAYTNTTYNSNWSNCGNACNGAY 291
QY 199 AAACAAGATCTAATTAATCTGTTTGGACACTGCATCCAAATTTTGGAACTGAGTTTCAGT 258
DB 292 CARTAYGAYTATACNTGYTNGARMGNGNWSNCAATYATYTGARGAARTAYWSN 351
QY 259 AAGTACTGCCAGCTGTTGCTGCTCTCTGTTGCTGAGATATCTCGAACAATCTCTCAT 318
DB 352 AARTYTYTGYCCNGCGNGTGYMGNGAYATHGCGNGGNGAYATHWSNNGNAAACNAARGAY 411
QY 319 GATATAGAGATCTCTGCCATTTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 378
DB 412 GNTATMGNGAYACNWSNYTNTGYAARGCNGCNATHCAYCGNCGNATHATHACNGAY 471
QY 379 AGTTTGGGGCGCAATCACTAGTGTGTAATTAATTAAGTAAAGTATTCCTTATTAAGATTTCT 438
DB 472 GARYTNGGNGNCAATYATHAAYTNTNCARWSNAARNGNATHWSNCAATYAGARGNGYTN 531
QY 439 TTGGCTAAACAACGCTACATCTGCTGCTGGACACTTATCTACAAAGTCTTTTACATTTAAG 498
DB 532 YTGCGNAAYGNGTNTNWSNMGNCAYGNGWSNYTWSNNGARAARMGNTTYTNTTYACN 591
QY 499 ACAAGTGGATGTATGGAACACTGGGGATGAGTCTGCTGCTGATC 543
DB 592 ACNCCNGGNATGAAYATHACNACNGTNGCNATHCCNWSNNGTNAH 636

RESULT 11
US-10-003-132-5
; Sequence 5, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 2868
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (110)...(1486)
US-10-003-132-5

Query Match 4.7%; Score 92.6; DB 9; Length 2868;
Best Local Similarity 53.8%; Pred. No. 2e-17;
Matches 191; Conservative 0; Mismatches 164; Indels 0; Gaps 0;
QY 189 TGTATAGATAAAACAAGATCTAATTAATCTGTTGGACACTGCATCCAAATTTTGGAAACC 248
DB 391 TGCAACAGATCAGTATGATTTAATAACCTGTTGGACGAGGACGCCATTATTTCCGAGGA 450
QY 249 TGAGTTTCAAGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 308
DB 451 AAAATACAGCAAAATTTCTGCCAGCTGCTGTAGAGACATAGCAGGAGATATTTCTGGAA 510
QY 309 AATTCCTGATATAGAGATTCCTGCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 368
DB 511 TACAAAAAGATGGTTACAGAGATACCTCTTATTTGTGCAAAAGCTGCCATCCAGCGGAT 570

Pred. No.: 20.3 Length: 231
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 17 Gaps: 0

US-10-060-830-4 (1-300) x Q976P5 (1-231)

QY 281 AGCGTCTCTTTTCCTCTCTG 258
 |||||
 Db 3 SerLeuLeuPheLeuLeu 10

RESULT 13

Q94CK4 PRELIMINARY; PRT; 257 AA.
 ID Q94CK4
 AC Q94CK4
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Annexin-like protein.
 GN AT5G12380.
 OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
 RA Rudd S., Schoof H., Mayer K.F.X.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL592312; CAC42899.1;
 DR InterPro: IPR001464; Annexin.
 DR Pfam: PF00191; annexin; 3.
 DR ProDom: PD000143; Annexin; 3.
 DR ProSITE: PS00223; ANNEXIN; UNKNOWN_1.
 SQ SEQUENCE 257 AA; 29013 MW; 4006DD3164EACB9A CRC64;

Alignment Scores:
 Pred. No.: 20.1 Length: 257
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.08% Indels: 0
 DB: 10 Gaps: 0

US-10-060-830-4 (1-300) x Q94CK4 (1-257)

QY 92 AAAGCAGTGCACCATGAGAAAC 115
 |||||
 Db 122 LysAlaValAspHisGluGluThr 129

RESULT 14

O69884 PRELIMINARY; PRT; 259 AA.
 ID O69884
 AC O69884
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative signal peptidase I.
 GN SIP1 OR SIPW OR SCO5596 OR SC2E1.13.
 OS Streptomyces coelicolor, and
 OS Streptomyces lividans.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomyces.
 OX NCBI_TaxID=1902, 1916;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TK21;
 RA Parro V., Mellado R.P.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.coelicolor; STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL023797; CAA19388.1;
 DR EMBL; 286111; CAB06807.1;
 DR HSP; P00803; 1b12.
 DR InterPro: IPR000508; SigPase.
 DR InterPro: IPR000223; SigPase_S26A.
 DR Pfam: PF00461; Peptidase_S26; 2.
 DR PRINTS: PR00727; LEADERPTASE.
 DR PROSITE: PS00761; SPASE_1_3; UNKNOWN_1.
 SQ SEQUENCE 259 AA; 27661 MW; 9D30961300A64A86 CRC64;

Alignment Scores:
 Pred. No.: 20.1 Length: 259
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.25% Indels: 0
 DB: 16 Gaps: 0

US-10-060-830-4 (1-300) x O69884 (1-259)

QY 278 CTGCTCTTTTCCTCTCTGTT 255
 |||||
 Db 56 LeuLeuPheLeuLeuLeu 63

RESULT 15

O9JUT4 PRELIMINARY; PRT; 260 AA.
 ID O9JUT4
 AC O9JUT4
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE GDNF family receptor alpha 4, GPI-anchored isoform.
 GN GPR4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=THYROID;
RX MEDLINE=20319126; PubMed=10860579;
RA Lindahl M., Rimmusk T., Rossi J., Saarma M., Airaksinen M.S.;
RT "Expression and alternative splicing of mouse Gfra4 suggest roles in
  endocrine cell development.";
RL Mol. Cell. Neurosci. 15:522-533(2000).
DR EMBL; AJ276870; CAB89690.1; -.
DR MGD; MGI:1341873; Gfra4.
DR InterPro; IPR003438; GDNF_receptor.
DR Pfam; PF02351; GDNF; 1.
DR PRINTS; PRO1316; GDNFRECEPTOR.
KW Receptor.
SQ SEQUENCE 260 AA; 27990 MW; 2679BBC789E38075 CRC64;

Alignment Scores:
Pred. No.:      20.1      Length:      260
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      8.08%      Indels:      0
DB:              11      Gaps:      0

US-10-060-830-4 (1-300) x Q9JTT4 (1-260)
QY 271 AGAAGCAGGCTATGCAGACCTAGA 294
  |||||
Db 132 ArgSerArgLeucysArgProArg 139
```

Search completed: January 21, 2003, 14:52:21
Job time : 33.9449 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 21, 2003, 14:46:52 ; Search time 10.8661 Seconds
(without alignments)
1624.657 Million cell updates/sec

Title: US-10-060-830-4

Perfect score: 99

Sequence: 1 aaacagaaagaaaaaactg.....tatgcagacatagatcctta 300

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Xgapop 60.0 , Xgapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438908

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q-/cgn2_1/USPTO_SPOOL/US10060830/runat_16012003_092703_1523/app_query.fasta_1.1109
-DB-IssuedPatents_AA -QFWI=fastan -SUFFIX=olin2p.ra -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10060830_ECGN_1.1.8 @runat_16012003_092703_1523 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-FGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A.COMB.pcp:*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pcp:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pcp:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pcp:*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pcp:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	8	8.2	448	4	US-09-310-463-18
C 2	8	8.2	448	4	US-08-842-248A-18
C 3	7	7.1	9	2	US-08-725-736D-18
C 4	7	7.1	9	3	US-09-162-368B-18
C 5	7	7.1	9	4	US-09-161-877B-18
C 6	7	7.1	10	2	US-08-725-736D-5
C 7	7	7.2	10	3	US-08-159-339A-488
C 8	7	7.1	10	3	US-09-162-368B-5
C 9	7	7.1	10	4	US-09-161-877B-5
C 10	7	7.1	10	5	PCT-US91-08328-25
C 11	7	7.1	11	5	PCT-US91-08328-15
C 12	7	7.2	25	1	US-07-966-187-4

C 13	7	7.2	25	1	US-08-371-121-3	Sequence 3, Appli
C 14	7	7.2	25	3	US-07-927-391-5	Sequence 5, Appli
C 15	7	7.2	30	4	US-09-262-856A-15	Sequence 15, Appli
C 16	7	7.2	174	1	US-08-261-825-2	Sequence 2, Appli
C 17	7	7.2	174	2	US-08-719-124-2	Sequence 2, Appli
C 18	7	7.2	174	5	PCT-US95-07748A-2	Sequence 2, Appli
C 19	7	7.2	226	4	US-09-134-001C-5590	Sequence 5590, Ap
C 20	7	7.2	226	3	US-09-400-742-2	Sequence 2, Appli
C 21	7	7.2	283	4	US-08-618-651A-2	Sequence 2, Appli
C 22	7	7.2	283	4	US-09-215-252-2	Sequence 2, Appli
C 23	7	7.2	340	1	US-08-324-977-34	Sequence 34, Appli
C 24	7	7.2	340	2	US-08-384-616-34	Sequence 34, Appli
C 25	7	7.2	340	2	US-08-904-686A-34	Sequence 34, Appli
C 26	7	7.2	340	4	US-09-315-850-34	Sequence 34, Appli
C 27	7	7.1	360	4	US-09-116-498-10	Sequence 10, Appli
C 28	7	7.1	360	4	US-09-116-498-12	Sequence 12, Appli
C 29	7	7.2	402	1	US-08-460-806-2	Sequence 2, Appli
C 30	7	7.2	402	1	US-08-460-806-4	Sequence 4, Appli
C 31	7	7.2	402	1	US-08-460-806-5	Sequence 5, Appli
C 32	7	7.2	402	1	US-08-460-806-7	Sequence 7, Appli
C 33	7	7.2	402	1	US-08-460-806-9	Sequence 9, Appli
C 34	7	7.2	402	1	US-08-460-806-11	Sequence 11, Appli
C 35	7	7.2	402	1	US-08-325-630-2	Sequence 2, Appli
C 36	7	7.2	402	1	US-08-325-630-4	Sequence 4, Appli
C 37	7	7.2	402	1	US-08-325-630-5	Sequence 5, Appli
C 38	7	7.2	402	1	US-08-325-630-7	Sequence 7, Appli
C 39	7	7.2	402	1	US-08-325-630-9	Sequence 9, Appli
C 40	7	7.2	402	1	US-08-325-630-11	Sequence 11, Appli
C 41	7	7.1	440	4	US-08-759-638-4	Sequence 4, Appli
C 42	7	7.2	463	4	US-08-612-973-46	Sequence 46, Appli
C 43	7	7.2	463	4	US-08-927-597-46	Sequence 46, Appli
C 44	7	7.2	480	1	US-08-440-103-18	Sequence 18, Appli
C 45	7	7.2	480	1	US-08-440-103-22	Sequence 22, Appli

ALIGNMENTS

RESULT 1

US-09-310-463-18
; Sequence 18, Application US/09310463A
; Patent No. 6384203
; GENERAL INFORMATION:
; APPLICANT: Cosman, David J.
; APPLICANT: Anderson, Dirk M.
; APPLICANT: Borges, Luis
; TITLE OF INVENTION: Family of Immunoregulators Designated Leukocyte Immunoglob
; TITLE OF INVENTION: Like Receptors (LIR)
; FILE REFERENCE: 2624-A
; CURRENT APPLICATION NUMBER: US/09/310,463A
; CURRENT FILING DATE: 1999-05-12
; EARLIER APPLICATION NUMBER: 08/842,248
; EARLIER FILING DATE: 1997-04-24
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 448
; TYPE: PRT
; ORGANISM: human
US-09-310-463-18

Alignment Scores:	4.22	Length:	448
Pred. No.:	8.00	Matches:	8
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	8.25%	Gaps:	0
DB:	4		
US-10-060-830-4 (1-300) x US-09-310-463-18 (1-448)			
Qy 281 AGCGTGTCTCTTTCTCTCTG 258			
Db 273 SerLeuLeuPheLeuLeu 280			

```
RESULT 2
US-08-842-248A-18
: Sequence 18, Application US/08842248A
: Patent No. 6448035
: GENERAL INFORMATION:
: APPLICANT: Cosman, David J.
: TITLE OF INVENTION: Family of Immunoregulators Designated
: TITLE OF INVENTION: Leukocyte Immunoglobulin-Like Receptors (LIR)
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Janis C. Henry, Immunex Corporation
: STREET: 51 University Street
: CITY: Seattle
: STATE: WA
: COUNTRY: US
: ZIP: 98101
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM/PC Compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/842,248A
: FILING DATE: April 24, 1997
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Henry, Janis C.
: REGISTRATION NUMBER: 34,347
: REFERENCE/DOCKET NUMBER: 2624
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 587-0430
: TELEFAX: (206) 233-0644
: TELEX: 756822
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 448 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-842-248A-18
Alignment Scores:
Pred. No.: 4.22 Length: 448
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.25% Indels: 0
DB: 4 Gaps: 0
US-10-060-830-4 (1-300) x US-08-842-248A-18 (1-448)
QY 281 AGCCTGCTTCTTTCTCTTCTG 258
|||||
Db 273 SerLeuLeuPheLeuLeu 280
RESULT 3
US-08-725-736D-18
: Sequence 18, Application US/08725736D
: Patent No. 5831016
: GENERAL INFORMATION:
: APPLICANT: WANG, R.F.; ROSENBERG, S. A.
: TITLE OF INVENTION: IDENTIFICATION OF TRP-2 AS
: TITLE OF INVENTION: A HUMAN TUMOR ANTIGEN RECOGNIZED BY CYTOTOXIC T
: TITLE OF INVENTION: LYMPHOCYTES
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
: STREET: 345 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/162,368B
FILING DATE: 09-FEB-1996
NAME: DOROTHY R. AUTH
ATTORNEY/AGENT INFORMATION:
REGISTRATION NUMBER: 36,434
REFERENCE/DOCKET NUMBER: 2026-4243
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 9
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
MOLECULE TYPE:
DESCRIPTION: PEPTIDE
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
US-08-725-736D-18
Alignment Scores:
Pred. No.: 6.43e+06 Length: 9
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 2 Gaps: 0
US-10-060-830-4 (1-300) x US-08-725-736D-18 (1-9)
QY 37 ACCTTACTGGGACCGGCAGG 57
|||||
Db 3 ThrLeuLeuGlyProGlyArg 9
RESULT 4
US-09-162-368B-18
: Sequence 18, Application US/09162368B
: Patent No. 6083703
: GENERAL INFORMATION:
: APPLICANT: WANG, R.F.; ROSENBERG, S. A.
: TITLE OF INVENTION: IDENTIFICATION OF TRP-2 AS
: TITLE OF INVENTION: A HUMAN TUMOR ANTIGEN RECOGNIZED BY CYTOTOXIC T
: TITLE OF INVENTION: LYMPHOCYTES
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
: STREET: 345 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/162,368B
```

;; FILING DATE: 28-SEPT-1998
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/725,736
;; FILING DATE: 04-OCT-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/599,602
;; FILING DATE: 09-FEB-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KATHRYN M. BROWN
;; REGISTRATION NUMBER: 34,556
;; REFERENCE/DOCKET NUMBER: 2026-4243US1
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;; TELEX: 421792
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 9
;; TYPE: AMINO ACID
;; STRANDEDNESS: UNKNOWN
;; TOPOLOGY: UNKNOWN
;; MOLECULE TYPE:
;; DESCRIPTION: PEPTIDE
;; FEATURE:
;; NAME/KEY:
;; LOCATION:
;; IDENTIFICATION METHOD:
;; OTHER INFORMATION:
US-09-162-368B-18

Alignment Scores:
Pred. No.: 6.43e+06 Length: 9
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 3 Gaps: 0

US-10-060-830-4 (1-300) x US-09-162-368B-18 (1-9)
QY 37 ACCTTACTGGACGGCAGG 57
Db 3 ThrLeuLeuGlyProGlyArg 9

RESULT 5
US-09-161-877B-18
; Sequence 18, Application US/09161877B
; Patent No. 6132980
; GENERAL INFORMATION:
; APPLICANT: WANG, R.F.; ROSENBERG, S. A.
; TITLE OF INVENTION: IDENTIFICATION OF TRP-2
; TITLE OF INVENTION: AS A HUMAN TUMOR ANTIGEN RECOGNIZED BY
; TITLE OF INVENTION: CYTOTOXIC T LYMPHOCYTES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/161,877B
; FILING DATE: 28-SEPT-1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/725,736

;; FILING DATE: 04-OCT-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/599,602
;; FILING DATE: 09-FEB-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KATHRYN M. BROWN
;; REGISTRATION NUMBER: 34,556
;; REFERENCE/DOCKET NUMBER: 2026-4243US2
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;; TELEX: 421792
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 9
;; TYPE: AMINO ACID
;; STRANDEDNESS: UNKNOWN
;; TOPOLOGY: UNKNOWN
;; MOLECULE TYPE:
;; DESCRIPTION: PEPTIDE
;; FEATURE:
;; NAME/KEY:
;; LOCATION:
;; IDENTIFICATION METHOD:
;; OTHER INFORMATION:
US-09-161-877B-18

Alignment Scores:
Pred. No.: 6.43e+06 Length: 9
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-4 (1-300) x US-09-161-877B-18 (1-9)
QY 37 ACCTTACTGGACGGCAGG 57
Db 3 ThrLeuLeuGlyProGlyArg 9

RESULT 6
US-08-725-736D-5
; Sequence 5, Application US/08725736D
; Patent No. 5831016
; GENERAL INFORMATION:
; APPLICANT: WANG, R.F.; ROSENBERG, S. A.
; TITLE OF INVENTION: IDENTIFICATION OF TRP-2 AS
; TITLE OF INVENTION: A HUMAN TUMOR ANTIGEN RECOGNIZED BY CYTOTOXIC T
; TITLE OF INVENTION: LYMPHOCYTES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,736D
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,602
; FILING DATE: 09-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36,434

```

; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (415) 576-0200
; ; TELEFAX: (415) 576-0300
; ; TELEX:
; ; INFORMATION FOR SEQ ID NO: 488:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 10 amino acids
; ; TYPE: amino acid
; ; STRANDEDNESS: single
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: peptide
; ; US-08-159-339A-488
; ;
; ; Alignment Scores:
; ; Pred. No.: 70.9
; ; Score: 7.00
; ; Percent Similarity: 100.00%
; ; Best Local Similarity: 100.00%
; ; Query Match: 7.22%
; ; DB: 3
; ;
; ; US-10-060-830-4 (1-300) x US-08-159-339A-488
; ;
; ; QY 278 CTGCTCTTTTCCTCTCTG 258
; ; |||||
; ; Db 4 LeuLeuLeuPheLeuLeu 10
; ;
; ; RESULT 8
; ; US-09-162-368B-5
; ; Sequence 5, Application US/09162368B
; ; Patent No. 6083703
; ; GENERAL INFORMATION:
; ; APPLICANT: WANG, R.F.; ROSENBERG, S. A.
; ; TITLE OF INVENTION: IDENTIFICATION OF T
; ; TITLE OF INVENTION: A HUMAN TUMOR ANTIG
; ; TITLE OF INVENTION: LYMPHOCYTES
; ; NUMBER OF SEQUENCES: 31
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; ; CITY: 345 PARK AVENUE
; ; CITY: NEW YORK
; ; STATE: NEW YORK
; ; COUNTRY: USA
; ; ZIP: 10154
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: FLOPPY DISK
; ; COMPUTER: IBM PC COMPATIBLE
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: MICROSOFT WORD 97
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/09/162,368B
; ; FILING DATE: 28-SEPT-1998
; ; CLASSIFICATION: 530
; ; PRIOR APPLICATION DATA:
; ; APPLICATION NUMBER: 08/725,736
; ; FILING DATE: 04-OCT-1996
; ; PRIOR APPLICATION DATA:
; ; APPLICATION NUMBER: 08/599,602
; ; FILING DATE: 09-FEB-1996
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: KATHRYN M. BROWN
; ; REGISTRATION NUMBER: 34,556
; ; REFERENCE/DOCKET NUMBER: 2076-4243051
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (212) 758-4800
; ; TELEFAX: (212) 751-6849
; ; TELEX: 421792
; ; INFORMATION FOR SEQ ID NO: 5:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 10
; ; TYPE: AMINO ACID
; ; STRANDEDNESS: SINGLE
; ; TOPOLOGY: UNKNOWN

```

```
; MOLECULE TYPE:
; DESCRIPTION: PEPTIDE
; FEATURE:
; NAME/KEY: TRP-2 PEPTIDE
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
US-09-162-368B-5

Alignment Scores:
Pred. No.: 70.9 Length: 10
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 3 Gaps: 0

US-10-060-830-4 (1-300) x US-09-162-368B-5 (1-10)
QY 37 ACCTTACTGGACCGGCAGG 57
Db 1 ThrLeuLeuGlyProGlyArg 7

RESULT 9
US-09-161-877B-5
; Sequence 5, Application US/09161877B
; Patent No. 6132980
; GENERAL INFORMATION:
; APPLICANT: WANG, R.F.; ROSENBERG, S. A.
; TITLE OF INVENTION: IDENTIFICATION OF TRP-2
; TITLE OF INVENTION: AS A HUMAN TUMOR ANTIGEN RECOGNIZED BY
; TITLE OF INVENTION: CYTOTOXIC T LYMPHOCYTES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/161,877B
; FILING DATE: 28-SEPT-1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/725,736
; FILING DATE: 04-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,602
; FILING DATE: 09-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: KATHRYN M. BROWN
; REGISTRATION NUMBER: 34,556
; REFERENCE/DOCKET NUMBER: 2026-42430S2
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: UNKNOWN
; MOLECULE TYPE:
; DESCRIPTION: PEPTIDE
; FEATURE:
; NAME/KEY: TRP-2 PEPTIDE
```

```
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
US-09-161-877B-5

Alignment Scores:
Pred. No.: 70.9 Length: 10
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 4 Gaps: 0

US-10-060-830-4 (1-300) x US-09-161-877B-5 (1-10)
QY 37 ACCTTACTGGACCGGCAGG 57
Db 1 ThrLeuLeuGlyProGlyArg 7

RESULT 10
PCT-US91-08328-25
; Sequence 25, Application PC/TUS9108328
; GENERAL INFORMATION:
; APPLICANT: Ruggeri, Zaverio M.
; APPLICANT: Houghten, Richard A.
; TITLE OF INVENTION: PEPTIDES THAT INHIBIT PLATELET BINDING
; TITLE OF INVENTION: OF ADHESION MOLECULES
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eugene Moroz, MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/08328
; FILING DATE: 19911107
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/610,363
; FILING DATE: 07-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Moroz, Eugene
; REGISTRATION NUMBER: 25,237
; REFERENCE/DOCKET NUMBER: 1198 4079PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: Cross-links
; LOCATION: 4..>10
; OTHER INFORMATION: /note= "Sequence linked by
; OTHER INFORMATION: Interchain amide bond at Lys residue with Glu
; OTHER INFORMATION: residue on Arg3-Glu-Arg-Ser-Arg-Gly-Asp-Val"
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: US 4,683,291
; FILING DATE: 28-OCT-1985
; PUBLICATION DATE: 28-JUL-1987
```

```
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: US B1 4,683,291
; FILING DATE: 28-OCT-1985
; PUBLICATION DATE: 03-JUL-1990
PCT-US91-08328-25

Alignment Scores:
Pred. No.: 70.9 Length: 10
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 5 Gaps: 0

US-10-060-830-4 (1-300) x PCT-US91-08328-25 (1-10)
QY 259 AGAAGAAGGAAGAGACAGG 279
Db 1 ArgArgArgLysArgSerArg 7

RESULT 11
; Sequence 15, Application PC/TUS9108328
; GENERAL INFORMATION:
; APPLICANT: Ruggeri, Zaverio M.
; TITLE OF INVENTION: PEPTIDES THAT INHIBIT PLATELET BINDING
; TITLE OF INVENTION: OF ADHESION MOLECULES
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eugene Moroz, MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/08328
; FILING DATE: 19911107
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/610,363
; FILING DATE: 07-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Moroz, Eugene
; REGISTRATION NUMBER: 25,237
; REFERENCE/DOCKET NUMBER: 1198 4079PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: Cross-links
; LOCATION: 5..>11
; OTHER INFORMATION: /note= Sequence linked by
; OTHER INFORMATION: Interchain amide bond at Lys residue with Glu
; OTHER INFORMATION: residue on Arg4-Glu-Arg-Ser-Arg-Gly-Asp-Val
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: US 4,683,291
; FILING DATE: 28-OCT-1985

; PUBLICATION DATE: 28-JUL-1987
; DOCUMENT NUMBER: US B1 4,683,291
; FILING DATE: 28-OCT-1985
; PUBLICATION DATE: 03-JUL-1990
PCT-US91-08328-15

Alignment Scores:
Pred. No.: 70 Length: 11
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 5 Gaps: 0

US-10-060-830-4 (1-300) x PCT-US91-08328-15 (1-11)
QY 259 AGAAGAAGGAAGAGACAGG 279
Db 2 ArgArgArgLysArgSerArg 8

RESULT 12
; Sequence 4, Application US/07966187
; Patent No. 5477001
; GENERAL INFORMATION:
; APPLICANT: SASS, Catherine
; APPLICANT: LEGUAY, Jean-Jacques
; APPLICANT: GRISON, Rene
; APPLICANT: TOPPAN, Alain
; TITLE OF INVENTION: Recombinant DNA coding for a novel
; TITLE OF INVENTION: protein having beta 1,3 glucanase activity, bacteria
; TITLE OF INVENTION: containing this DNA, transformed plant cells and plants.
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: King Street Station, Suite 500, 1800 Diagonal
; STREET: Road, PO Box 299
; CITY: ALEXANDRIA
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/966,187
; FILING DATE: 19930125
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR92/00268
; FILING DATE: 25-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91/03588
; FILING DATE: 25-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: SAXE, Bernhard D
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 16781/381
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-966-187-4
```

Alignment Scores:
Pred. No.: 62.3 Length: 25
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x US-07-966-187-4 (1-25)

QY 278 CTGCTCTCTTCTCTCTCTG 258
DB 10 LeuLeuLeuPheLeuLeuLeu 16

RESULT 13

US-08-371-121-3

; Sequence 3, Application US/08371121

; Patent No. 5652123

; GENERAL INFORMATION:

; APPLICANT: CAPUT, Daniel

; APPLICANT: FERRARA, Pascual

; APPLICANT: GUILLEMOT, Jean-Claude

; APPLICANT: LEPLATOIS, Pascal

; APPLICANT: MINTY, Adrian

; APPLICANT: KAGHAD, Mourad

; APPLICANT: LABIT-LE BOUTEILLER, Christine

; APPLICANT: MAGAZIN, Marilyn

; TITLE OF INVENTION: Protein having a cytokine type

; TITLE OF INVENTION: activity, recombinant DNA coding for this protein,

; TITLE OF INVENTION: transformed cells and microorganisms.

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FOLEY & LARDNER

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington, D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/371,121

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/938,161

; FILING DATE: 30-NOV-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/FR92/00280

; FILING DATE: 27-MAR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 91 00137

; FILING DATE: 08-JAN-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 91 03904

; FILING DATE: 29-MAR-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: SAXE, Bernhard D.

; REGISTRATION NUMBER: 28,665

; REFERENCE/DOCKET NUMBER: 16781/383

; TELEPHONE: (202) 672-5300

; TELEFAX: (202) 672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 25 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-371-121-3

Alignment Scores:
Pred. No.: 62.3 Length: 25
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 1 Gaps: 0

US-10-060-830-4 (1-300) x US-08-371-121-3 (1-25)

QY 278 CTGCTCTCTTCTCTCTCTG 258
DB 10 LeuLeuLeuPheLeuLeuLeu 16

RESULT 14

US-07-927-391-5

; Sequence 5, Application US/07927391

; Patent No. 6001649

; GENERAL INFORMATION:

; APPLICANT: CAPUT, Daniel

; APPLICANT: FERRARA, Pascual

; APPLICANT: MILOUX, Brigitte

; APPLICANT: MINTY, Adrian

; APPLICANT: VITA, Natalio

; TITLE OF INVENTION: Protein having a cytokin type

; TITLE OF INVENTION: activity, and recombinant DNA, expression vector and host

; TITLE OF INVENTION: for its preparation.

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FOLEY & LARDNER

; STREET: King Street Station, Suite 500, 1800 Diagonal

; CITY: ALEXANDRIA

; STATE: VIRGINIA

; COUNTRY: USA

; ZIP: 22313-0299

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/927,391

; FILING DATE: 19920929

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: SAXE, Bernhard D.

; REGISTRATION NUMBER: 28,665

; REFERENCE/DOCKET NUMBER: 16781/369

; TELEPHONE: (703) 836-9300

; TELEFAX: (703) 683-4109

; TELEX: 899149

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 25 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-07-927-391-5

Alignment Scores:
Pred. No.: 62.3 Length: 25
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.22% Indels: 0
DB: 3 Gaps: 0

US-10-060-830-4 (1-300) x US-07-927-391-5 (1-25)

QY 278 CTGCTCTCTTCTCTCTCTG 258

```
Db      10 LeuLeuPheLeuLeuLeu 16
|||||
RESULT 15
US-09-262-856A-15
; Sequence 15, Application US/09262856A
; Patent No. 6333164
; GENERAL INFORMATION:
; APPLICANT: TAKESAKO, Kazutoh
; APPLICANT: MIZUTANI, Shigetoshi
; APPLICANT: ENDO, Masahiro
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: FUNGAL ANTIGENS AND PROCESS FOR PRODUCING THE SAME
; FILE REFERENCE: 1422-372P
; CURRENT APPLICATION NUMBER: US/09/262,856A
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-262-856A-15

Alignment Scores:
Pred. No.:      60.7      Length:      30
Score:          7.00      Matches:      7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:      7.22% Indels: 0
DB:               A      Gaps: 0

US-10-060-830-4 (1-300) x US-09-262-856A-15 (1-30)

Qy      107 CATGGTCCACTGCTTTTGCAG 87
|||||
Db      24 HisGlyProLeuLeuLeuGln 30

Search completed: January 21, 2003, 14:53:19
Job time : 11.8661 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:49:45 ; Search time 13.2301 Seconds
(without alignments)
4744.920 Million cell updates/sec

Title: US-10-060-830-1114

Perfect score: 3470

Sequence: 1 MFLFLLLVLLVLLLEEDAGA.....TQEVSGAGRGDCDFVKEIL 653

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	361	10.4	2211	1 KFB05	coagulation factor
2	355.5	10.2	2183	2 T42764	coagulation factor
3	351	10.1	2324	1 KPHU5	coagulation factor
4	348	10.0	401	2 S65138	glycoprotein anti
5	348	10.0	427	2 S74211	PAS-6/7 protein pr
6	343	9.9	218	2 A47285	milk fat globule p
7	339	9.8	409	2 T11743	PP47 protein - pig
8	334	9.6	427	2 JC4915	ags protein precu
9	321	9.3	463	1 A36479	milk fat globule m
10	308.5	8.9	216	2 A44258	factor VIII-associ
11	308.5	8.9	2319	2 A47004	coagulation factor
12	308.5	8.9	2351	1 EZHU	coagulation factor
13	307	8.8	2133	2 T42763	coagulation factor
14	258	7.4	927	1 JQ0948	A5 antigen precurs
15	224.5	6.5	845	2 JC5256	adipocyte transcr
16	213	6.1	3133	2 S52093	hemocytin - silkw
17	207.5	6.0	1283	2 T13799	neurexin IV - fru
18	204	5.9	719	2 S51739	transcription repr
19	200.5	5.8	855	2 S42621	protein-tyrosine k
20	187	5.4	913	2 A48280	receptor tyrosine
21	179	5.2	910	2 A53137	tyrosine kinase re
22	171	4.9	819	2 I48859	tyro 10 receptor k
23	167	4.8	876	2 A49508	protein-tyrosine k
24	145	4.2	1381	2 T31083	paranodin - rat
25	145	4.2	1385	2 T14158	neurexin IV - mous
26	136	3.9	737	2 T15615	hypothetical prote
27	127	3.7	737	2 T31349	hypothetical prote
28	127	3.7	837	1 A29512	LDL receptor precu
29	126.5	3.6	854	1 QRHYLD	LDL receptor precu

30 124.5 3.6 3570 2 T45025
31 123.5 3.6 1019 2 A38738
32 122.5 3.5 1843 2 S18803
33 122.5 3.5 4006 2 T09070
34 122 3.5 996 2 J02037
35 121.5 3.5 678 2 T49984
36 120.5 3.5 760 2 F86387
37 120.5 3.5 862 1 ORMSLD
38 120 3.5 3020 2 A43932
39 119.5 3.4 699 2 I38073
40 119 3.4 1140 2 D88690
41 117.5 3.4 1827 2 T34288
42 117 3.4 514 2 H70699
43 116 3.3 414 2 S30397
44 116 3.3 1075 2 S54067
45 115.5 3.3 1427 2 I51669

ALIGNMENTS

RESULT 1

KFB05

coagulation factor V precursor - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999

C:Accession: A42580; A36497

R:Guinto, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.

J. Biol. Chem. 267, 2971-2978, 1992

A:Title: The complete cDNA sequence of bovine coagulation factor V.

A:Reference number: A42580; MUID:92147638; PMID:1737753

A:Accession: A42580

A:Molecule type: mRNA

A:Residues: 1-2211 <GUI>

A:Cross-references: GB:M81440; NID:g163037; PIDN:AAA30512.1; PID:g163038

A:Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBI:P:80776)

R:Kalafatis, M.; Jenny, R.J.; Mann, K.G.

J. Biol. Chem. 265, 21580-21589, 1990

A:Title: Identification and characterization of a phospholipid-binding site of bovi

A:Reference number: A36497; MUID:91072354; PMID:2254316

A:Accession: A36497

A:Molecule type: Protein

A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584, 1673-1676, 'X', 1678-1679, 'X', 1681,

R:Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.

Biochemistry 33, 13109-13116, 1994

A:Title: Determination of the disulfide bridges in factor Va heavy chain.

A:Reference number: A55979; MUID:95034740; PMID:7947716

A:Contents: annotation

A:Note: 566-Cys and 617-Cys were shown to have free sulphydryls

C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa

C:Function:

A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa

C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxida:

C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; pla:

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-2211/Product: coagulation factor V #status predicted <MAT>

F:29-741/Product: coagulation factor Va heavy chain #status predicted <VAH>

F:29-345/Domain: A1 <DA1>

F:33-329/Domain: ferroxidase repeat homology <FO1>

F:346-695/Domain: A2 <DA2>

F:351-688/Domain: ferroxidase repeat homology <FO2>

F:696-1564/Domain: B <DOB>

F:1175-1437/Region: 9-residue repeats (O-X-T/N-L-S-P-D-L-S)

F:1565-1892/Domain: A3 <DA3>

F:1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>

F:1572-1892/Domain: ferroxidase repeat homology <FO3>

F:1654-1752/Region: phospholipid binding #status predicted

F:1893-2051/Domain: C1 <DC1>

F:1893-2048/Domain: discoidin I amino-terminal homology <DN1>

F:2052-2211/Domain: C2 <DC2>

F:2052-2208/Domain: discoidin I amino-terminal homology <DN2>

F:167-193,248-329,499-525/Disulfide bonds: #status experimental

F:225,239,297,382,460,553,587,745,756,774,780,902,952,964,1044,1053,1062,1071,1078,1094,
F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:363,697,1537/Binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:533-534/Cleavage site: Arg-Gly (protein C) #status predicted
F:607-688,1712-1738,1894-2048,2053-2208/Disulfide bonds: #status predicted
F:741-742/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F:1034-1035/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F:1564-1565/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 10.4%; Score 361; DB 1; Length 2211;
Best Local Similarity 47.6%; Pred. No. 2,3e-17;
Matches 79; Conservative 17; Mismatches 60; Indels 10; Gaps 3;

QY 166 KTSYCYGTGLMESGYTADPQITASSVLE--WDHTQENSWPKKRLKPG--PWAAP 221
DB 2049 EYVNGSTPLGMSGKTKENKQITASSPKKSW-----GNYWEPFLARLNAAQGRVNAWQAK 2102
QY 222 AYDEYQWLQIDLNKKEKIGITGTSTWYEHNVYSAIRLYSDGQKVTYRPGVEOD 281
DB 2103 ANNNQWLQIDLLKIKITAIYVQGCSSLSEMYKSYIHYSDQGTDKWPKREKSSWD 2162
QY 282 KTFQGNKDYHÖVRNFFLPPIIARFIRVNPTQWQKIAMKMLGCG 327
DB 2163 KTFEGNNVRGHVKNFNPPIISRFIRIIPKTNQSIARLRLFGC 2208

RESULT 2
T42764
coagulation factor V - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42764
R:Yang, T.L.; Cui, J.; Rehmulla, A.; Yang, A.; Moussalli, M.; Kaufman, R.J.; Ginsburg,
Blood 91, 4593-4599, 1998
A:Title: The structure and function of murine factor V and its inactivation by protein C
A:Reference number: 22270; MUID:98282202; PMID:9616155
A:Accession: T42764
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2183 <YAN>
A:Cross-references: EMBL:U52925; NID:g3219690; PID:g3219691; PIDN:AAC99553.1
C:Function:
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
F:350-682/Domain: ferroxidase repeat homology <FOX1>
F:1541-1864/Domain: ferroxidase repeat homology <FOX2>

Query Match 10.2%; Score 355.5; DB 2; Length 2183;
Best Local Similarity 40.1%; Pred. No. 5.8e-17;
Matches 79; Conservative 28; Mismatches 75; Indels 15; Gaps 4;

QY 135 VISKGIPIYESSANNVTSVGHLSLFTFTKTSYGTGLMESGVITADPQITASSVLE- 193
DB 1995 IVARYIRIHTPKSYNRP-----LRLEQCEVNGCGSTPLGLEDRQDKQTASSPKKS 2049
QY 194 -WDHTQENSWPKKRLKPG--PWAFAFDEYQWLQIDLNKKEKIGITGTSTW 250
DB 2050 WNGDY-----WEPFLARLNAAQGRVNAWQAKANNKQWLQVLLKIKKTAIVTQGCSSL 2103
QY 251 EHNYSAYIRLYSDGQKVTYRPGVEODKIFQGNKDYHQDVRRNFFLPPIIARFIRVN 310
DB 2104 SSEMYKSYIQYSDQVANKPYRQKSSWDKIFEGNNTKGMKNFNPPIISRFIRI 2163
QY 311 PTWQKQIAMKMLGCG 327
DB 2164 PKTNQSIARLRLFGC 2180

RESULT 3
KFH05
coagulation factor V precursor [validated] - human

N:Alternate names: coagulation labile factor; proaccelerin
C:Species: Homo sapiens (man)
C:Date: 19-May-1989 #sequence_revision 02-Jun-1995 #text_change 08-Dec-2000
C:Accession: A56172; A23344; A28028; A27498; A25897
R:Cripe, L.D.; Moore, K.D.; Kane, W.H.
Biochemistry 31, 3777-3785, 1992
A:Title: Structure of the gene for human coagulation factor V.
A:Reference number: A42344; MUID:92232668; PMID:1567832
A:Accession: A56172
A:Molecule type: DNA
A:Residues: 1-2224 <CRI>
A:Cross-references: GB:J05368
A:Accession: A42344
A:Molecule type: DNA
A:Residues: 48-58; 79-89; 120-130; 191-201; 239-249; 313-323; 368-378; 428-437; 461-471; 533-5
2070; 2111-2120; 2172-2181 <CR2>
R:Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Ka
Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
A:Title: Complete cDNA and derived amino acid sequence of human factor V.
A:Reference number: A28028; MUID:87260886; PMID:3110773
A:Accession: A28028
A:Molecule type: mRNA
A:Residues: 1-857; R', 859-864, R', 866-924, E', 926-1763, I', 1765-2212, T', 2214-2224 <J
A:Cross-references: GB:M16967
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.
Biochemistry 26, 6508-6514, 1987
A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of
A:Reference number: A27498; MUID:88107560; PMID:2827731
A:Accession: A27498
A:Molecule type: mRNA
A:Residues: 1-1284, I', 1286-1600 <KAN>
A:Cross-references: GB:M17785
A:Note: parts of this sequence were determined by protein sequencing
R:Kane, W.H.; Davie, E.W.
Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homo
A:Reference number: A25897; MUID:86313665; PMID:3092220
A:Accession: A25897
A:Molecule type: mRNA
A:Residues: 1188-1215, 1315-2224 <KA2>
A:Cross-references: GB:M14335
A:Note: parts of this sequence were determined by protein sequencing
R:Keller, F.G.; Ortel, T.L.; Quinn-Allen, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A:Title: Thrombin-catalyzed activation of recombinant human factor V.
A:Reference number: A56139; MUID:95210278; PMID:7696276
A:Contents: annotation; thrombin cleavage sites
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Genetics:
A:Gene: GDB:F5
A:Cross-references: GDB:119896; OMIM:227400
A:Map position: 1q23-1q23
A:Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1;
C:Function:
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasm
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-2224/Product: coagulation factor V #status predicted <MAT>
F:29-737/Product: coagulation factor Va heavy chain #status experimental <VAH>
F:33-329/Domain: ferroxidase repeat homology <FO1>
F:346-691/Domain: A2 <DA2>
F:351-684/Domain: ferroxidase repeat homology <FO2>
F:692-1573/Domain: B <DOB>
F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F:1574-1905/Domain: A3 <DA3>
F:1581-1905/Domain: ferroxidase repeat homology <FO3>
F:1667-1765/Region: phospholipid binding #status predicted
F:1906-2064/Domain: C1 <DC1>

```

      | | :||||| |:||||||: | ||::||: | ||| || | |
288 KFNWTAQTNSASEWLQIDLGSQRKVTGIIITOGARDFGHIQVAAIRVAYGDGVTTWEY 347

      QY   274 RFGVEOQIKFQGKNKYHQDVNRNFPPIIARFI RVNPNTQMOOKIAMKMELGGC 327
           ::|| | ||| ||: | | -||-||: | | :||| |||
Db    348 KDPCASSEKFIPGNNDNNSHKKNFETPQAREFVRIQPVANHNRITLRVELLGC 401


RESULT 5
S74211
PAS-6/7 protein precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
C:Accession: S74211; S78114; S24181
R:Ivarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A:Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat
A:Reference number: S74211; MUID:97008954; PMID:8856064
A:Accession: S74211
A:Molecule type: mRNA
A:A:Residues: 1-427 <HVA>
A:A:Cross-references: EMBL:X91895; NID:g1632778; PIDN:CAA62997.1; PID:g1632779
A:A:Accession: S78114
A:Molecule type: protein
A:A:Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;422-425
Biochim. Biophys. Acta 1122, 203-211, 1992
A:Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7,
A:Reference number: S23926; MUID:92353107;PMID:1643094
```

A:Residues: 383-394 <KIM>

[illegible]

RESULT 6

A47285
 milk fat globule protein - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
 C:Accession: A47285
 R:Tarocca, D.; Peterson, J.A.; Urrea, R.; Kuniyoshi, J.; Bistrain, A.M.; Ceriani, R.
 Cancer Res. 51, 4994-4998, 1991
 A:Title: A M-r 46,000 human milk fat globule protein that is highly expressed in hu
 A:Reference number: A47285; PMID:91371351; PMID:1909932


```

Db 2176 SIRSTLRMELMGCDLNSCSMPLGMSKASLSDAQITASSYF-----TNMFATWSPSKARLH 2230
QY 213 KPG--PPWAAAFATDEYQWLQIDLNEKKKITGIITGSTWVEHNYVVSAYRILYSDDGOKW 270
Db 2231 LGRSNAWRPVNNPKWQLQVDFOKTKMKTGYTTGGVSKLTSMYVKEFLISSQDGHQW 2250
QY 271 TVYREGVEODKIFQGNKHQDVDRNNFLPPLIARIPIRVNPTQWQOKIAMKMWLLGCO 328
Db 2291 TLPFQNG--KKVYFOGNQDSFTPVVNSLDPLLLTRYLRIHQSWWYHQIALRMEVLGCE 2346

RESULT 13
T42763
coagulation factor VIII precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42763
R:Lollar, P.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z22269
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Accession: T42763
A:Molecule type: mRNA
A:Residues: 1-2133 <LOI>
A:Cross-references: EMBL:U49517; NID:g1511633; PID:g1511634; PIDN:AAB06705,
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology
C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hem
F:1-19/Domain: signal sequence status predicted <Sig>
F:2-2133/Product: coagulation factor VIII #status predicted <MAT>
F:23-349/Domain: ferroxidase repeat homology <FOX1>
F:402-730/Domain: ferroxidase repeat homology <FOX2>
F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query Match 8.8%; Score 307; DB 2; Length 2133;
Best Local Similarity 26.3%; Pred.No. 1.8e-13;
Matches 110; Conservative 51; Mismatches 163; Indels 94; Gaps

QY 44 TLLFMSGIHVSGRGLASYSVIDKQDLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGT 103
Db 1663 TLNAAHGROVTVQEFALFTFD-----ETKSWYFTVENRCRAPHQLQMED--PT 1712
QY 104 IPHGIRDSPPLCMAGVHA--GVVSNFLGQISVLSKIGPIYESLANN----- 150
Db 1713 LKENYR-----FHAINGVYMDPLGLV-MAQNQRIRYLLSMGNSENIHSIHFSGH 1762
QY 151 -----VTSVVG-----HLSTSL-----FTFKTSGC 170
Db 1763 VFSVRKKEEKMAVYNLYPCVFETVEMLPKVGIVRIECLIGEHLQAGMSTFLVYSKEC 18222
QY 171 YGTLGMSGVADPQITASVLEWDHTGQENSMPKPKARLKPKGPWAAFPATDEYQWLQ 230
Db 1823 QAPLGMASGRIDFQITAS-----GQYGOWAPKARLHRYSGSINAWSTKDPHSWIK 1873
QY 231 IDLNKKKKTGIITGSTWVEHNYVVSAYRILYSDGOKWTVYREGVEQDKIFQGNKDY 290
Db 1874 VDLLAPMIHGIWTQAROKFSSLIQIPIMISLDGRNWSYRGNSTGTLMVFGNDA 1933
QY 291 HQDVRRNNFLPIIARIPIRVNPTQWQOKIAMKMWLLGCOF-----IPKGRPPKLTQPPPPRN 346
Db 1934 SGIKHNFPPIVARIYRLHPHYHSIRSLRMELMGCDLNSCSMPLGMQNKAISDSQITA 1993
QY 347 SNDLKN--TTAPPKIAKGRAPKFTQPLQPRRSNEFPATQETTSPTDTRNTVTPNVT 402
Db 1994 SSSLNIFATWSPSQARLHQGRNTAWNRPRVS-----SABEWLQVDLOKTVKVTGIT 2045

```

RESULT 14
JQ0948
A5 antigen precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: JH0466; JQ0948
R:Takagi, S.; Hirata, T.; Agata, K.; Mochii, M.; Emauchi, G.; Fujiisawa, H.

```

A:Cross-references: DDBJ: D86479; NID:g1468942; PIDN:BAAL3094.1; PID:g1468943

Query Match      6.5%; Score 224.5; DB 2; Length 845;
Best Local Similarity 31.1%; Pred. No. 4.6e-08;
Matches 52; Conservative 28; Mismatches 68; Indels 19; Gaps 4;

QY 174 LGMESGVADPQITASSVLEWTDHTGQBSNWKPKKARLKKP-----GPPWAAAFATD 224
      :||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75 IGMESHRIEDNQIIRASNL-----RHGLGAQGRGLNMQTGATEDDYYDGAWCAEDDA 126

QY 225 EYQWLQIDLNKEKITGIITGTSWMEYHNYYSAYRILYSDGQKWTYVREPGVQDKIF 284
      |||:|:|:|:| | | | | | | | | | | | | | | | | | | | | | | | |
Db 127 RTQWIEVDTRTRFTGTGITQGRDSSIHDFVTFVFGFSNDQSOTWVMYTN-GYE-EMTF 184

QY 285 QGNKDYHODYRNNLPPITFIARFIRVNPQWQOKIAMKELLGCGPI 331
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 185 HGNVDKOTPVLSLPPEPVPVAFIRIYPLTWNGSICMRLEVLGCSVAP 231

Search completed: January 21, 2003, 09:54:43
Job time : 21.2301 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:52:40 ; Search time 8.03256 Seconds
(without alignments)
3371.781 Million cell updates/sec

Title: US-10-060-830-1114

Perfect score: 3470

Sequence: 1 MPLFLLLLVLLLEDAGA.....TQEVSGAGRGCECDVFKEL 653

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	362	10.4	2258	1	FA5_PIG	Q9glp1 sus scrofa
2	361	10.4	2211	1	FA5_BOVIN	P21077 bos taurus
3	348	10.0	427	1	MFGM_BOVIN	Q95114 bos taurus
4	347	10.0	2224	1	FA5_HUMAN	P12259 homo sapien
5	343	9.9	387	1	MFGM_HUMAN	Q08431 homo sapien
6	339	9.8	409	1	MFGM_PIG	P79385 sus scrofa
7	334	9.6	427	1	MFGM_RAT	P70490 rattus norv
8	321	9.3	463	1	MFGM_MOUSE	P21956 mus musculus
9	308.5	8.9	2319	1	FA8_MOUSE	Q06194 mus musculus
10	308.5	8.9	2351	1	FA8_MOUSE	P00451 homo sapien
11	307	8.8	2133	1	FA8_PIG	P12263 sus scrofa
12	295	8.5	925	1	NRP2_RAT	Q35276 rattus norv
13	295	8.5	931	1	NRP2_HUMAN	O60462 homo sapien
14	294	8.5	931	1	NRP2_MOUSE	O35375 mus musculus
15	284.5	8.2	914	1	NRP1_CHICK	P79795 gallus gall
16	271	7.8	922	1	NRP1_RAT	Q9qwj9 rattus norv
17	271	7.8	923	1	NRP1_MOUSE	P97333 mus musculus
18	269.5	7.8	224	1	XLRI_MOUSE	Q92114 mus musculus
19	265	7.6	923	1	NRP1_HUMAN	O14786 homo sapien
20	261.5	7.5	224	1	XLRI_HUMAN	O15537 homo sapien
21	258	7.4	928	1	NRP1_XENLA	P28824 xenopus lae
22	256.5	7.4	280	1	XLRI_FUGRU	Q9uwr5 fugu rubrip
23	217	6.3	1331	1	CTA2_HUMAN	Q9uhc6 homo sapien
24	213	6.1	3133	1	HMCT_BOMMO	P98092 bombyx mori
25	208	6.0	1310	1	CTA4_MOUSE	Q99p47 mus musculus
26	207.5	6.0	1284	1	NRX4_DROME	Q94887 drosophila
27	206	5.9	1308	1	CTA4_HUMAN	Q9cua0 homo sapien
28	200.5	5.8	855	1	DDR2_HUMAN	Q16832 homo sapien
29	186	5.4	913	1	DDR1_HUMAN	Q08345 h epithelia
30	184.5	5.3	911	1	DDR1_MOUSE	O03146 mus musculus
31	184	5.3	854	1	DDR2_MOUSE	O62371 mus musculus
32	182	5.2	1288	1	CTA3_HUMAN	Q9bz76 homo sapien
33	179	5.2	910	1	DDR1_RAT	O63474 rattus norv

34	150	4.3	1384	1	CTAL_HUMAN	P78357 homo sapien
35	145	4.2	1381	1	CTAL_RAT	P97846 rattus norv
36	145	4.2	1385	1	CTAL_MOUSE	O54991 mus musculus
37	136	3.9	257	1	DIS2_DICDI	P42530 dictyosteli
38	135	3.9	867	1	SSPO_BOVIN	P98167 bos taurus
39	127	3.7	837	1	LDLR_RABIT	P20063 oryctolagus
40	126.5	3.6	854	1	LDLR_CRIGR	P35950 cricetus
41	123.5	3.6	1019	1	LFC_TACTR	Q28175 tachyples
42	120	3.5	5179	1	MUC2_HUMAN	Q02817 homo sapien
43	119.5	3.4	699	1	NP14_HUMAN	O14978 homo sapien
44	119.5	3.4	1019	1	LFC_CARRO	Q28422 carcinoscor
45	117.5	3.4	864	1	LDLR_MOUSE	P35951 mus musculus

ALIGNMENTS

RESULT 1

FA5_PIG STANDARD; PRT; 2258 AA.
AC Q9GLP1:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Sulna; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C DOMAINS.
RC TISSUE=Liver;
RA MEDLINE=21121490; PubMed=11229814;
RA Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,
RA Kim H.K.W.;
RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional protein modeling of membrane binding sites and comparative anatomy of domains.";
RT Cell. Mol. Life Sci. 58:148-159(2001).
RL CC
CC -!- FUNCTION: Coagulation factor V is a cofactor that participates with factor Xa to activate prothrombin to thrombin.
CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light chain, noncovalently bound. The interaction between the two chains is calcium-dependent.
CC -!- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1 and C2 may be involved in membrane binding.
CC -!- PTM: Thrombin activates factor V proteolytically to the active cofactor, factor Va (formation of a heavy chain at the N-terminus and a light chain at the C-terminus).
CC -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF 2 PLASTOCYANIN-LIKE REPEATS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF191308; AAG28381.1; --
HSSP: P12259; 1CZT
InterPro: IPR001117; Cu-oxidase.
InterPro: IPR000421; FA58_C.
Pfam: PF00394; Cu-oxidase; 3.
Pfam: PF00754; F5_F8_type_C; 2.
SMART: SM00231; FA58C; 2.
PROSITE: PS00079; MULTICOPPER_OXIDASE1; 2.
PROSITE: PS01285; FA58C_1; 2.

DR PROSITE: PS01286; FA58C.2; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 2258 COAGULATION FACTOR V.
FT CHAIN 23 737 HEAVY CHAIN (BY SIMILARITY).
FT PEPTIDE 738 1611 ACTIVATION PEPTIDE (CONNECTING REGION)
(BY SIMILARITY).
FT CHAIN 1612 2258 LIGHT CHAIN (BY SIMILARITY).
FT DOMAIN 30 329 F5/8 TYPE A 1.
FT DOMAIN 30 193 PLASTOCYANIN-LIKE 1.
FT DOMAIN 203 329 PLASTOCYANIN-LIKE 2.
FT DOMAIN 348 683 F5/8 TYPE A 2.
FT DOMAIN 348 525 PLASTOCYANIN-LIKE 3.
FT DOMAIN 535 683 PLASTOCYANIN-LIKE 4.
FT DOMAIN 691 1611
FT DOMAIN 1168 1539
FT REPEAT 1168 1176
FT REPEAT 1177 1185
FT REPEAT 1186 1194
FT REPEAT 1195 1203
FT REPEAT 1204 1212
FT REPEAT 1213 1221
FT REPEAT 1222 1230
FT REPEAT 1231 1239
FT REPEAT 1240 1248
FT REPEAT 1249 1257
FT REPEAT 1258 1266
FT REPEAT 1267 1275
FT REPEAT 1276 1284
FT REPEAT 1285 1293
FT REPEAT 1294 1302
FT REPEAT 1303 1311
FT REPEAT 1312 1320
FT REPEAT 1321 1329
FT REPEAT 1330 1338
FT REPEAT 1339 1347
FT REPEAT 1348 1356
FT REPEAT 1357 1365
FT REPEAT 1366 1374
FT REPEAT 1375 1383
FT REPEAT 1384 1392
FT REPEAT 1393 1401
FT REPEAT 1402 1410
FT REPEAT 1411 1419
FT REPEAT 1420 1428
FT REPEAT 1429 1437
FT REPEAT 1438 1446
FT REPEAT 1447 1455
FT REPEAT 1456 1464
FT REPEAT 1465 1473
FT REPEAT 1474 1482
FT REPEAT 1483 1491
FT REPEAT 1492 1500
FT REPEAT 1501 1509
FT REPEAT 1510 1518
FT REPEAT 1519 1527
FT REPEAT 1531 1539
FT DOMAIN 1616 1941 F5/8 TYPE A 3.
FT DOMAIN 1616 1785 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1795 1941 PLASTOCYANIN-LIKE 6.
FT DOMAIN 1942 2095
FT DOMAIN 2100 2255 F5/8 TYPE C 1.
FT SITE 737 738 F5/8 TYPE C 2.
FT SITE 1029 1030 CLEAVAGE (BY THROMBIN) (BY
SIMILARITY).
FT SITE 1611 1612 CLEAVAGE (BY THROMBIN) (BY
SIMILARITY).
FT DISULFID 167 193 PROBABLE.
FT DISULFID 499 525 PROBABLE.
FT DISULFID 1759 1785 PROBABLE.

FT DISULFID 1941 2095 BY SIMILARITY.
FT DISULFID 2100 2255 BY SIMILARITY.
FT MOD_RES 692 696 SULFATION (POTENTIAL).
FT MOD_RES 696 696 SULFATION (POTENTIAL).
FT MOD_RES 724 724 SULFATION (POTENTIAL).
FT MOD_RES 726 726 SULFATION (POTENTIAL).
FT MOD_RES 745 745 SULFATION (POTENTIAL).
FT MOD_RES 1560 1560 SULFATION (POTENTIAL).
FT MOD_RES 1576 1576 SULFATION (POTENTIAL).
FT MOD_RES 1581 1581 SULFATION (POTENTIAL).
FT MOD_RES 1584 1584 SULFATION (POTENTIAL).
FT MOD_RES 1588 1588 SULFATION (POTENTIAL).
FT MOD_RES 1631 1631 SULFATION (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 460 460 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 467 467 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 553 553 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 741 741 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 752 752 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 760 760 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 776 776 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 960 960 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1048 1048 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1057 1057 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1066 1066 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1480 1480 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1537 1537 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1597 1597 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1886 1886 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2243 2243 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2258 AA; 256078 MW; 9159B9E0076A2ACC CRC64;
Query Match 10.4%; Score 362; DB 1; Length 2258;
Best Local Similarity 47.6%; Pred. No. 2.5e-17;
Matches 79; Conservative 18; Mismatches 59; Indels 10; Gaps 3;
Qy 166 KTSQCYGTGLMESGVIAADPQITASSVLE--WTDHTGQENSWKPKRKARLKKPG--PPWAAF 221
Db 2096 EVNGCSTPLGMESGNIKNEQITASSFKKSWGDY-----WEFFARLNAQGRVNAQAK 2149
Qy 222 ATDEYQWLOIDLNKKKTIITGTSTWVEHNYYSAYRILYSDGQKWTVTYREPGEVD 281
Db 2150 ANNNQWLQIDLLKIKKITAITQCKSLSEMYVKRYTQISDRGVENKYSREKSSMYD 2209
Qy 282 KIFQGNKDYHODVRNFPPIIARFIRVPTQWQKIAMKMLGCG 327
Db 2210 KIFEGNNNIKGVKNFPPIISIRFIIRPKMWNOSIALRLEFGC 2255
RESULT 2
FA5_BOVIN STANDARD; PRT: 2211 AA.
ID FA5_BOVIN Q28107; Q28108;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.

NCBI_TaxID=9913;
 [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Liver;
 RC MEDLINE=92147638; PubMed=1737753;
 RX Guinto E.R., Esmon C.T., Mann K.G., Macgillivray R.T.;
 RA "The complete cDNA sequence of bovine coagulation factor V.";
 RL J. Biol. Chem. 267:2971-2978(1992).
 CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -1- DOMAIN: Domain B contains 29.5 x 9 AA tandem repeats, and 2 x 14
 CC AA repeats.
 CC -1- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
 CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
 CC 2 PLASTOCYANIN-LIKE REPEATS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M61440; AAA30512.1; -;
 DR EMBL: M61441; AAA30513.1; -;
 DR HSSP: P12259; IC2T.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR000421; FA58_C.
 DR Pfam: PF00394; Cu-oxidase; 3.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR SMART: SM00231; FA58C; 2.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 2.
 DR PROSITE: PS01285; FA58C.1; 2.
 DR PROSITE: PS01286; FA58C.2; 2.
 DR Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
 KW Repeat.
 KW SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 2211 COAGULATION FACTOR V.
 FT CHAIN 29 741 HEAVY CHAIN (BY SIMILARITY).
 FT PEPTIDE 742 1564 ACTIVATION PEPTIDE (CONNECTING REGION)
 FT (BY SIMILARITY).
 FT CHAIN 1565 2211 LIGHT CHAIN (BY SIMILARITY).
 FT DOMAIN 30 327 F5/8 TYPE A 1.
 FT DOMAIN 30 193 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 203 327 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 348 686 F5/8 TYPE A 2.
 FT DOMAIN 348 525 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 535 686 PLASTOCYANIN-LIKE 4.
 FT DOMAIN 696 1564 B.
 FT SIMILAR 899 915 TO 17 AA REPEATS IN HUMAN FA5.
 FT DOMAIN 1124 1151 2 x 14 AA TANDEM REPEATS.
 FT REPEAT 1124 1137 1-1.
 FT REPEAT 1138 1151 1-2.
 FT DOMAIN 1188 1453 30 x 9 AA APPROXIMATE TANDEM REPEATS, OF
 FT [AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
 FT REPEAT 1188 1196 2-1.
 FT REPEAT 1197 1205 2-2.
 FT REPEAT 1206 1214 2-3.
 FT REPEAT 1215 1223 2-4.
 FT REPEAT 1224 1232 2-5.
 FT REPEAT 1233 1241 2-6.
 FT REPEAT 1242 1250 2-7.
 FT REPEAT 1251 1259 2-8.

FT REPEAT 1268 1268 2-9.
 FT REPEAT 1269 1268 2-10.
 FT REPEAT 1278 1286 2-11.
 FT REPEAT 1287 1295 2-12.
 FT REPEAT 1296 1304 2-13.
 FT REPEAT 1305 1313 2-14.
 FT REPEAT 1314 1322 2-15.
 FT REPEAT 1323 1331 2-16.
 FT REPEAT 1332 1340 2-17.
 FT REPEAT 1341 1349 2-18.
 FT REPEAT 1350 1358 2-19.
 FT REPEAT 1359 1367 2-20.
 FT REPEAT 1368 1376 2-21.
 FT REPEAT 1377 1385 2-22.
 FT REPEAT 1386 1394 2-23.
 FT REPEAT 1395 1403 2-24.
 FT REPEAT 1404 1412 2-25.
 FT REPEAT 1413 1421 2-26.
 FT REPEAT 1422 1430 2-27.
 FT REPEAT 1431 1439 2-28.
 FT REPEAT 1440 1444 2-29 (PARTIAL).
 FT REPEAT 1445 1453 2-30.
 FT DOMAIN 1569 1890 F5/8 TYPE A 3.
 FT DOMAIN 1748 1890 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 1894 2048 PLASTOCYANIN-LIKE 6.
 FT DOMAIN 2053 2208 F5/8 TYPE C 1.
 FT SITE 741 742 F5/8 TYPE C 2.
 FT SITE 1034 1035 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT SITE 1564 1565 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
 FT DISULFID 167 193 PROBABLE.
 FT DISULFID 499 525 PROBABLE.
 FT DISULFID 1712 1738 BY SIMILARITY.
 FT DISULFID 1894 2048 BY SIMILARITY.
 FT MOD_RES 697 697 SULFATION (POTENTIAL).
 FT MOD_RES 701 701 SULFATION (POTENTIAL).
 FT MOD_RES 730 730 SULFATION (POTENTIAL).
 FT MOD_RES 1513 1513 SULFATION (POTENTIAL).
 FT MOD_RES 1529 1529 SULFATION (POTENTIAL).
 FT MOD_RES 1537 1537 SULFATION (POTENTIAL).
 FT MOD_RES 1541 1541 SULFATION (POTENTIAL).
 FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 902 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 952 952 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1094 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1490 1490 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1690 1690 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1839 1839 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1997 1997 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2196 2196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 587 592 N*TLPA -> T (IN VARIANT 2).
 SQ SEQUENCE 2211 AA; 248981 MW; CBBF90B738667C45 CRC64;

10.4%; Score 361; DB 1; Length 2211;

Query Match

RESULT 4
FAS_HUMAN
ID FAS_HUMAN STANDARD; PRT; 2224 AA.
AC P12259; Q14285;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92232668; PubMed=1567832;
RA Crispe L.D., Moore K.D., Kane W.H.;
RT "Structure of the gene for human coagulation factor V.";
RL Biochemistry 31:3777-3785(1992).
[2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=87260886; PubMed=3110773;
RA Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
Hewick R.M., Kaufman R.J., Mann K.G.;
RT "Complete cDNA and derived amino acid sequence of human factor V.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
[3]
RN RP SEQUENCE OF 1-1600 FROM N.A.
RX MEDLINE=88107560; PubMed=2827731;
RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
RT "Cloning of cDNAs coding for the heavy chain region and connecting
region of human factor V, a blood coagulation factor with four types
of internal repeats.";
RL Biochemistry 26:6508-6514(1987).
[4]
RN RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
RX MEDLINE=86313665; PubMed=3092220;
RA Kane W.H., Davie E.W.;
RT "Cloning of a cDNA coding for human factor V, a blood coagulation
factor homologous to factor VIII and ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
[5]
RN RP PARTIAL SEQUENCE FROM N.A.
RC TISSUE-Fibroblast;
RX MEDLINE=93203619; PubMed=8454869;
RA Shen N.L.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,
Edgington T.S.;
RT "The serine protease cofactor factor V is synthesized by
lymphocytes.";
RL J. Immunol. 150:2992-3001(1993).
[6]
RN RP SULFATION.
RX MEDLINE=94264012; PubMed=8204629;
RA Pittman D.D., Tomkinson K.N., Michnick D., Seligsohn U.,
Kaufman R.J.;
RT "Posttranslational sulfation of factor V is required for efficient
thrombin cleavage and activation and for full procoagulant activity.";
RL Biochemistry 33:6952-6959(1994).
[7]
RN RP SULFATION.
RX MEDLINE=90366699; PubMed=2168225;
RA Hortic G.L.;
RT "Sulfation of tyrosine residues in coagulation factor V.";
RL Blood 76:946-952(1990).
[8]
RN RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
RX MEDLINE=2052169; PubMed=10586886;
RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
Fuentes-Prior P.;
RT "Crystal structures of the membrane-binding C2 domain of human
coagulation factor V.";
RL Nature 402:434-439(1999).

RN RP VARIANT APCR GLN-534.
RX MEDLINE=94217810; PubMed=8164741;
RA Bertina R.M., Koelmen B.P.C., Koster T., Rosendaal F.R.,
Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
RT "Mutation in blood coagulation factor V associated with resistance to
activated protein C.";
RL Nature 369:64-67(1994).
[10]
RN RP VARIANT APCR GLY-334, AND VARIANT LYS-513.
RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese.";
RL Blood 91:1135-1139(1998).
[11]
RN RP VARIANT APCR THR-334.
RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
resistance to activated protein C.";
RL Blood 91:1140-1144(1998).
CC -!- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
chain, noncovalently bound. The interaction between the two chains
is calcium-dependent.
CC -!- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
REPEATS.
CC -!- PTM: Thrombin activates factor V proteolytically to the active
terminus, factor Va (formation of a heavy chain at the N-
terminus and a light chain at the C-terminus).
CC -!- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
CC -!- DISEASE: OWREN PARAHEMOPHILIA, AN HEMORRHAGIC DIATHESIS, IS DUE
TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APCR
MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
IMPLANTATION.
CC -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
2 PLASTOCYANIN-LIKE REPEATS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; L32779; AAB59401.1; JOINED.
DR EMBL; L32755; AAB59401.1; JOINED.
DR EMBL; L32756; AAB59401.1; JOINED.
DR EMBL; L32757; AAB59401.1; JOINED.
DR EMBL; L32758; AAB59401.1; JOINED.
DR EMBL; L32759; AAB59401.1; JOINED.
DR EMBL; L32760; AAB59401.1; JOINED.
DR EMBL; L32761; AAB59401.1; JOINED.
DR EMBL; L32762; AAB59401.1; JOINED.
DR EMBL; L32763; AAB59401.1; JOINED.
DR EMBL; L32764; AAB59401.1; JOINED.
DR EMBL; L32765; AAB59401.1; JOINED.
DR EMBL; L32766; AAB59401.1; JOINED.
DR EMBL; L32767; AAB59401.1; JOINED.
DR EMBL; L32768; AAB59401.1; JOINED.
DR EMBL; L32769; AAB59401.1; JOINED.
DR EMBL; L32770; AAB59401.1; JOINED.
DR EMBL; L32771; AAB59401.1; JOINED.
DR EMBL; L32772; AAB59401.1; JOINED.
DR EMBL; L32773; AAB59401.1; JOINED.
DR EMBL; L32774; AAB59401.1; JOINED.
DR EMBL; L32775; AAB59401.1; JOINED.
DR EMBL; L32776; AAB59401.1; JOINED.

```
DR EMBL: L32777; AAB59401.1; JOINED.
DR EMBL: L32778; AAB59401.1; JOINED.
DR EMBL: M16967; AAB52424.1; -.
DR EMBL: M14335; AAB59532.1; -.
DR PIR: A25897; A25897.
DR PIR: A28028; A28028.
DR PDB: 1CZS; 26-NOV-99.
DR PDB: 1CZT; 26-NOV-99.
DR PDB: 1CZV; 26-NOV-99.
DR Genew: HGNC:3542; F5.
DR MIM: 227400; -.
DR MIM: 134400; -.
DR MIM: 188055; -.
DR MIM: 227310; -.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00394; Cu-oxidase; 3.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR SMART: SM00231; FA58C; 2.
DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat; Polymorphism; Disease mutation; Thrombophilia; 3D-structure.
FT SIGNAL 1 28
FT CHAIN 29 2224 COAGULATION FACTOR V.
FT CHAIN 29 737 HEAVY CHAIN.
FT PEPTIDE 738 1573 ACTIVATION PEPTIDE (CONNECTING REGION).
FT CHAIN 1574 2224 LIGHT CHAIN.
FT DOMAIN 30 329 F5/8 TYPE A 1.
FT DOMAIN 30 193 PLASTOCYANIN-LIKE 1.
FT DOMAIN 203 329 PLASTOCYANIN-LIKE 2.
FT DOMAIN 348 684 F5/8 TYPE A 2.
FT DOMAIN 348 526 PLASTOCYANIN-LIKE 3.
FT DOMAIN 536 684 PLASTOCYANIN-LIKE 4.
FT DOMAIN 692 1573 B.
FT DOMAIN 895 928 2 X 17 AA TANDEM REPEATS.
FT REPEAT 895 911 1-1.
FT REPEAT 912 928 1-2.
FT SIMILAR 1135 1148 TO 14 AA REPEATS IN BOVINE FA5.
FT DOMAIN 1185 1501 35 X 9 AA APPROXIMATE TANDEM REPEATS OF
[TNP]-L-S-P-D-L-S-Q-T.
FT REPEAT 1185 1193 2-1.
FT REPEAT 1194 1202 2-2.
FT REPEAT 1203 1211 2-3.
FT REPEAT 1212 1220 2-4.
FT REPEAT 1221 1229 2-5.
FT REPEAT 1230 1238 2-6.
FT REPEAT 1239 1247 2-7.
FT REPEAT 1248 1256 2-8.
FT REPEAT 1257 1265 2-9.
FT REPEAT 1266 1274 2-10.
FT REPEAT 1275 1283 2-11.
FT REPEAT 1284 1292 2-12.
FT REPEAT 1293 1301 2-13.
FT REPEAT 1302 1310 2-14.

Query Match 10.08; Score 347; DB 1; Length 2224;
Best Local Similarity 45.8%; Pred. No. 2.8e-16;
Matches 76; Conservative 21; Mismatches 59; Indels 10; Gaps 3;

Qy 166 KTSQCYGLMGESVIADPOITASSVLE--WTDHTGOENSKPKKARKLKPG--PPWAAF 221
Db 2062 EVNGSTPLGMENKIEKNTQCKSLSEMYVKSTIHYSEGVKPKYRLKSSVD 2175
Qy 222 ATDEYQWLQDLNKKKTKITGTTGSTVHEHNYVSAIRYLSDGQKWTYVREPGVQD 281
Db 2116 ANNNKQWLIEDLLKTKITAITQCKSLSEMYVKSTIHYSEGVKPKYRLKSSVD 2175
Qy 282 KIFGKNKYDHDVNNFPLPIIARFIRVNTQWQOKIAMKMLLGC 327
Db 2176 KIFGNTKGVNFFNPPIISIRFIRVPTKWNQSTLRLELFGC 2221
```

RESULT 5

```
MFGM_HUMAN STANDARD; PRT; 387 AA.
ID MFGM_HUMAN
AC Q08431;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFG-E8
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=breast, and Breast carcinomas;
RX MEDLINE=96213908; PubMed=8639264;
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RT *Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain.*;
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
SEQUENCE OF 170-387 FROM N.A.
RP TISSUE=Mammary gland;
RX MEDLINE=91371351; PubMed=1905932;
RA Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
RA Ceriani R.L.;
RT *A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.*;
RL Cancer Res. 51:4994-4998(1991).
RN [3]
PARTIAL SEQUENCE, AND CHARACTERIZATION.
RP TISSUE=Milk;
RX MEDLINE=98194924; PubMed=9535276;
RA Guiffreda M.G., Cavaletto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT *Isolation and characterization of full and truncated forms of human
RT breast carcinoma protein BA46 from human milk fat globule membranes.*;
RL J. Protein Chem. 17:143-148(1998).
RN [4]
SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RX MEDLINE=99342076; PubMed=10411933;
RA Haegqvist B., Naeslund J., Sletten K., Westmark G.T., Mucchiano G.,
RA Tjernberg L.O., Nordstedt C., Engstrom U., Westmark P.;
RT *Medin: an integral fragment of aortic smooth muscle cell-produced
RT lactadherin forms the most common human amyloid.*;
RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
RN [5]
CHARACTERIZATION.
RX MEDLINE=97405885; PubMed=9260929;
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT *Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT (RGD)-dependent cell adhesion.*;
RL DNA Cell Biol. 16:861-869(1997).
CC -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
CC SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
CC -!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
CC MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC -!- PFM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
CC AMINO ACID 264 AND 273.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```



```

GN MFGE8 OR AGS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RC MEDLINE=96374422; PubMed=8780713;
RA Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
RT "Cloning and expression of cDNA for O-acetylation of GD3
RL ganglioside.";
RL Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO
CC PARTICIPATE IN THE O-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL; D84068; BAA12210.1; -.
DR HSPG; P00740; LEDM.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR PRINTS; PR00010; EGFBLOOD.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT SIGNAL 1 22
FT CHAIN 23 427
FT DOMAIN 24 61
FT DOMAIN 64 108
FT DOMAIN 111 267
FT DOMAIN 272 427
FT DISULFID 28 39
FT DISULFID 33 49
FT DISULFID 51 60
FT DISULFID 68 79
FT DISULFID 73 96
FT DISULFID 98 107
FT DISULFID 111 267
FT DISULFID 254 258
FT DISULFID 272 427
FT SITE 87 89
FT CARBOHYD 61 61
FT CARBOHYD 230 230
FT CARBOHYD 280 280
FT CARBOHYD 390 390
SQ SEQUENCE 427 AA; 94713 MW; EA8C8631F3EE6047 CRC64;
-----
Query Match 9.6%; Score 334; DB 1; Length 427;
Best Local Similarity 41.6%; Pred. No. 2.5e-16;
Matches 67; Conservative 27; Mismatches 61; Indels 6; Gaps 3;

QY 169 GCYTCLNMGESVIADPOITASSVLEWTDHTGCQNSWKPKRKALKKPGP--PWAPAFATDEY 226
|||::: ||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 271 GCSEPLGLKNNTIPDSOITASS--SYKTWNLRAPFGWYPHLGRCDNKGKINAWTAQSNAK 328

```

CC	-1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC	2 PLASTOCYANIN-LIKE REPEATS.
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC	-1- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC	or send an email to license@isb-sib.ch .
CC	-----
EMBL:	L05573; AAA37385.1; -
DR	PIR: A47004; A47004.
DR	HSSP: P00451; ICFG.
DR	MGG: MGI-88383; F8.
DR	InterPro: IPR001117; Cu-oxidase.
DR	InterPro: IPR000421; FA58_C.
DR	Pfam: PF00394; Cu-oxidase; 3.
DR	Pfam: PF00754; F5_F8_type_C; 2.
DR	SMART: SM00231; FA58C; 2.
DR	PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.
DR	PROSITE: PS01285; FA58C_1; 2.
DR	PROSITE: PS01286; FA58C_2; 2.
KW	Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
KW	Signal; Glycoprotein; Sulfation.
FT	SIGNAL 1 19 POTENTIAL.
FT	CHAIN 20 2319 COAGULATION FACTOR VIII.
FT	DOMAIN 20 349 F5/8 TYPE A 1.
FT	DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT	DOMAIN 207 349 PLASTOCYANIN-LIKE 2.
FT	DOMAIN 399 730 F5/8 TYPE A 2.
FT	DOMAIN 399 573 PLASTOCYANIN-LIKE 3.
FT	DOMAIN 583 730 PLASTOCYANIN-LIKE 4.
FT	DOMAIN 760 1640 B.
FT	DOMAIN 1683 2008 F5/8 TYPE A 3.
FT	DOMAIN 1683 1845 PLASTOCYANIN-LIKE 5.
FT	DOMAIN 1855 2008 PLASTOCYANIN-LIKE 6.
FT	DOMAIN 2009 2156 F5/8 TYPE C 1.
FT	DOMAIN 2161 2313 F5/8 TYPE C 2.
FT	SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	SITE 1678 1679 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	SITE 1324 1325 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT	SITE 1640 1641 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT	MOD_RES 367 367 SULFATION (BY SIMILARITY).
FT	MOD_RES 737 737 SULFATION (BY SIMILARITY).
FT	MOD_RES 738 738 SULFATION (BY SIMILARITY).
FT	MOD_RES 742 742 SULFATION (BY SIMILARITY).
FT	MOD_RES 1669 1669 SULFATION (REQUIRED FOR VWF BINDING)
FT	(BY SIMILARITY).
FT	MOD_RES 1687 1687 SULFATION (BY SIMILARITY).
FT	DISULFID 173 199 PROBABLE.
FT	DISULFID 547 573 PROBABLE.
FT	DISULFID 1819 1845 BY SIMILARITY.
FT	DISULFID 2008 2156 BY SIMILARITY.
FT	DISULFID 2161 2313 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 61 61 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 601 601 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 880 880 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 958 958 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1015 1015 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1022 1022 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1026 1026 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1044 1044 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1076 1076 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1087 1087 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1136 1136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1161 1161 N-LINKED (GLCNAC. .) (POTENTIAL).

[1] RP SEQUENCE FROM N.A.
RX MEDLINE-93265012; PubMed-1303178;
RA Gitschier J., Wood W.I.;
RT "Sequence of the exon-containing regions of the human factor VIII
RT gene.";
RL Hum. Mol. Genet. 1:199-200(1992).
[5] RP
RA SEQUENCE OF 2064-2070 FROM N.A.
de Water N.S., Williams R., Browett P.J.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
[6] RP
RN SULFATION OF TYR-1699.
RX MEDLINE-91093266; PubMed-1898735;
RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P.,
RA Mertens K., van Mourik J.A.;
RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
RT essential for the interaction of factor VIII with von Willebrand
RT factor.";
RL J. Biol. Chem. 266:740-746(1991).
[7] RP
RN SULFATION.
RX MEDLINE-95207952; PubMed-1554716;
RA Pittman D.D., Wang J.H., Kaufman R.J.;
RT "Identification and functional importance of tyrosine sulfate
RT residues within recombinant factor VIII.";
RL Biochemistry 31:3315-3325(1992).
[8] RP
RN STRUCTURE-BY NMR OF 2322-2343.
RX MEDLINE-95200924; PubMed-7893714;
RA Gilbert G.E., Balleja J.D.;
RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
RT amphipathic structure as determined by NMR spectroscopy.";
RL Biochemistry 34:3022-3031(1995).
[9] RP
RN REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE-91221499; PubMed-1902642;
RA Gitschier J.;
RT "The molecular basis of hemophilia A.";
RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
[10] RP
RN REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE-89088506; PubMed-2491949;
RA White G.C. II, Shoemaker C.B.;
RT "Factor VIII gene and hemophilia A.";
RL Blood 73:1-12(1989).
[11] RP
RN REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE-95245332; PubMed-7728145;
RA Antonarakis S.E., Kazanian H.H., Tuddenham E.G.D.;
RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
RL Hum. Mutat. 5:1-22(1995).
[12] RP
RN VARIANT GLN-2326.
RX MEDLINE-86235434; PubMed-3012775;
RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
RT "Identification of a missense mutation in the factor VIII gene of a
RT mild hemophilic.";
RL Science 232:1415-1416(1986).
[13] RP
RN VARIANT PRO-2135.
RX MEDLINE-88096539; PubMed-3122181;
RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
RT "A novel missense mutation in the factor VIII gene identified by
RT analysis of amplified hemophilia DNA sequences.";
RL Nucleic Acids Res. 15:9797-9805(1987).
[14] RP
RN VARIANT GLN-2228.
RX MEDLINE-88191889; PubMed-2833855;
RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
RA Kazanian H.H.;
RT "Nonsense and missense mutations in hemophilia A: estimate of the
RT relative mutation rate at CG dinucleotides.";

RL Am. J. Hum. Genet. 42:719-725(1988).
 RN [15]
 RP VARIANT GLY-291.
 RX MEDLINE-88220354; PubMed-2835904;
 RA Youssoufian H., Wong C., Aronis S., Platakoukis H., Kazazian H.H. Jr.,
 RA Antonarakis S.E.;
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
 in exon 7 of the factor VIII gene.";
 RL Am. J. Hum. Genet. 42:867-871(1988).
 RN [16]
 RP VARIANT CYS-1708.
 RX MEDLINE-89274393; PubMed-2499363;
 RA O'Brien D.P., Tuddenham E.G.;
 RT "Purification and characterization of factor VIII 1,689-Cys: a
 nonfunctional cofactor occurring in a patient with severe hemophilia
 A.";
 RL Blood 73:2117-2122(1989).
 RN [17]
 RP VARIANT CYS-391.
 RX MEDLINE-90001543; PubMed-2506948;
 RA Shima M., Ware J., Yoshioka A., Fukui H., Fulcher C.A.;
 RT "An arginine to cysteine amino acid substitution at a critical
 thrombin cleavage site in a dysfunctional factor VIII molecule.";
 RL Blood 74:1612-1617(1989).
 RN [18]
 RP VARIANT LEU-189.
 RX MEDLINE-90057680; PubMed-2510835;
 RA Chan V., Chan T.K., Tong T.M., Todd D.;
 RT "A novel missense mutation in exon 4 of the factor VIII:C gene
 resulting in moderately severe hemophilia A.";
 RL Blood 74:2688-2691(1989).
 RN [19]
 RP VARIANT LEU-2326.
 RX MEDLINE-89197216; PubMed-2495245;
 RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
 RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
 of the factor VIII gene.";
 RL Hum. Genet. 81:335-338(1989).
 RN [20]
 RP VARIANT HIS-391.
 RX MEDLINE-89264602; PubMed-2498882;
 RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Fujimaki M., Hoyer L.W.;
 RT "Direct characterization of factor VIII in plasma: detection of a
 mutation altering a thrombin cleavage site
 (arginine-372-->histidine).";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
 RN [21]
 RP VARIANT CYS-1708.
 RX MEDLINE-90105723; PubMed-2104766;
 RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
 Cys) in the factor VIII gene of two unrelated patients with
 cross-reacting material-positive hemophilia A.";
 RL Blood 75:384-389(1990).
 RN [22]
 RP VARIANTS GLN-2228 AND LEU-2326.
 RX MEDLINE-90123183; PubMed-2105106;
 RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
 RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;
 RT "Recurrent mutations and three novel rearrangements in the factor
 VIII gene of hemophilia A patients of Italian descent.";
 RL Blood 75:662-670(1990).
 RN [23]
 RP VARIANT CYS-391.
 RX MEDLINE-90329422; PubMed-1973901;
 RA Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
 RT "CRM+ haemophilia A due to a missense mutation (372-->Cys) at the
 internal heavy chain thrombin cleavage site.";
 RL Br. J. Haematol. 75:73-77(1990).
 RN [24]
 RP VARIANTS PHE-1699 AND CYS-1708.

RX MEDLINE-90152691; PubMed-2105906;
 RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K.,
 RA Kazazian H.H., Antonarakis S.E.;
 RT "Characterization of mutations in the factor VIII gene by direct
 sequencing of amplified genomic DNA.";
 RL Genomics 6:65-71(1990).
 RN [25]
 RP VARIANTS CYS-1728 AND ASP-1941.
 RX MEDLINE-90169988; PubMed-2106480;
 RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
 RA Kazazian H.H.;
 RT "Use of denaturing gradient gel electrophoresis to detect point
 mutations in the factor VIII gene.";
 RT mutations in the factor VIII gene.";
 Query Match 8.9%; Score 308.5; DB 1; Length 2351;
 Best Local Similarity 37.1%; Pred. No. 1.6e-13;
 Matches 66; Conservative 31; Mismatches 72; Indels 9; Gaps 3;
 QY 153 SVVGHLSLSTFTKTSCTGLMESGVADPOITASSVLEWTDHTGQENSKPKARLK 212
 DB 2176 SIRSLRMELMGCDLNSCSMPLGMSKASDAQITASSYF-----TNMFATWSPSKARLH 2230
 QY 213 KPG--PPWAAFAATDEYQWLQIDLNKKIKITITGTSTWVHNHYVSAYRILYSDGQKW 270
 DB 2231 LQGRSNARPPQVNNPKENLQVDFQTKMKTGVTGTVQGVKSLTSMYKFEFLISSQDGHQW 2290
 QY 271 TVYRPGVEQDKIFQGNKYHODVRNLFPIIARFIRVNPOTQOQKIAMKMLLGCO 328
 DB 2291 TLFFQNG--KVKVFGQNGDSFTPVNSLDPPLRLRYLRIPQSWVHQIALRMEVLGCE 2346
 RESULT 11
 ID FA8_PIG STANDARD; PRT; 2133 AA.
 AC P12263; Q95243;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Coagulation factor VIII precursor (procoagulant component).
 GN F8 OR CF8.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Healey J.F., Lubin I.M., Lollar P.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 705-1573 FROM N.A.
 RX MEDLINE-86287369; PubMed-3016730;
 RA Toole J.J., Pittman D.B., Orr E.C., Murtha P., Wasley L.C.,
 RA Kaufman R.J.;
 RT "A large region (approximately equal to 95 kDa) of human factor VIII
 is dispensable for in vitro procoagulant activity.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5939-5942(1986).
 RN [3]
 RP SEQUENCE OF 392-759 FROM N.A.
 RX MEDLINE-94179260; PubMed-7510693;
 RA Lubin I.M., Healey J.F., Scandella D., Runge M.S., Lollar P.;
 RT "Elimination of a major inhibitor epitope in factor VIII.";
 RL J. Biol. Chem. 269:8639-8641(1994).
 CC -1- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
 AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
 ACTIVATED FORM, FACTOR XA.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
 2 PLASTOCYANIN-LIKE REPEATS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U49517; AAB06705.1; -;
CC PIR; A25945; A25945.
CC HSP; P00451; ICFG.
CC InterPro; IPR001117; Cu-oxidase.
CC InterPro; IPR000421; FA58_C.
CC Pfam; PF00394; Cu-oxidase; 3.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00231; FA58C; 2.
CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
CC Signal; Glycoprotein; Sulfation.
CC SIGNAL 1 19 POTENTIAL.
CC CHAIN 20 2133 COAGULATION FACTOR VIII.
CC DOMAIN 20 357 F5/8 TYPE A 1.
CC DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
CC DOMAIN 207 357 PLASTOCYANIN-LIKE 2.
CC DOMAIN 399 730 F5/8 TYPE A 2.
CC DOMAIN 399 573 PLASTOCYANIN-LIKE 3.
CC DOMAIN 583 730 PLASTOCYANIN-LIKE 4.
CC DOMAIN 780 1599 B.
CC DOMAIN 1495 1822 F5/8 TYPE A 3.
CC DOMAIN 1495 1659 PLASTOCYANIN-LIKE 5.
CC DOMAIN 1669 1822 PLASTOCYANIN-LIKE 6.
CC DOMAIN 1823 1970 F5/8 TYPE C 1.
CC DOMAIN 1975 2127 F5/8 TYPE C 2.
CC SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
CC SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
CC SITE 1449 1450 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
CC SITE 1490 1491 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
CC MOD_RES 737 737 SULFATION (BY SIMILARITY).
CC MOD_RES 738 738 SULFATION (BY SIMILARITY).
CC MOD_RES 742 742 SULFATION (BY SIMILARITY).
CC DISULFID 173 199 PROBABLE.
CC DISULFID 547 573 PROBABLE.
CC DISULFID 1633 1659 PROBABLE.
CC DISULFID 1822 1970 BY SIMILARITY.
CC DISULFID 1975 2127 BY SIMILARITY.
CC CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 601 601 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 929 929 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 985 985 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1025 1025 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1111 1111 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1181 1181 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1208 1208 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1245 1245 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1265 1265 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1335 1335 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1408 1408 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1611 1611 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1919 1919 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CONFLICT 713 713 N -> M (IN REF. 2).
CC CONFLICT 734 734 I -> T (IN REF. 2).
CC CONFLICT 792 792 G -> Q (IN REF. 2).
CC CONFLICT 1133 1133 E -> F (IN REF. 2).
CC CONFLICT 1191 1191 I -> L (IN REF. 2).
CC CONFLICT 1209 1209 R -> F (IN REF. 2).
CC CONFLICT 1437 1437 C -> G (IN REF. 2).
CC CONFLICT 1456 1456 F -> R (IN REF. 2).
CC CONFLICT 1539 1539 Q -> R (IN REF. 2).
CC CONFLICT 1546 1546 Q -> N (IN REF. 2).
CC SEQUENCE 2133 AA; 239304 MW; 152BBA8997F570DA CRC64;
8.8%; Score 307; DB 1; Length 2133;

Query Match

Best Local Similarity 26.3%; Pred. No. 1.8e-13;
Matches 110; Conservative 51; Mismatches 163; Indels 94; Gaps 12;
QY 44 TLPMSGIHVGGRFLASYSVIDKDLITCDTASNFLEPFSKYCPAGCLLPFAEISGT 103
DB 1663 TLNAAHGRQVTQVEFALFTIFD-----ETKSWYFTENVERNCRAPCHQMED--PT 1712
QY 104 IPHGIRDSSPLCMAGVHA--GVSNLTGGQISVSVKSGIPYVESLANN----- 150
DB 1713 LKENYR-----FHAINGYVMDTLPGLV-MAQNRIRWYLLSMGNSNENIHHSFGH 1762
QY 151 -----VTSVVG-----HLSTSL---FTFKTSGC 170
DB 1763 VFSVRKKEEKYKAVNLYPGVFETVEMLPKVGIVRIECLIGEHLQAGMSTFLVYSKEC 1822
QY 171 YGTLMESGVIAIDPQITASSVLEWTDHTGQNSWKPKKARKLKFPWPAAAFATDEYOMLQ 230
DB 1823 QAPLMASGRIRDQITAS-----GOYGQWAPKLARLHYSGSINAWSTKDPHSHIK 1873
QY 231 IDLNKKEKITITGSTWVEHNVVSAIRILYSDGQKWTYREPGVEQDKIFQGNKDY 290
DB 1874 VDLLAPMIHIGIMTOGARQKFSLYISOFITMYSLDGRNWSYRGNSTGTLVWFFGNVDA 1933
QY 291 HQDVRRNFPPIIARFIRVNPTOMQOKIAMKELLGCOF-----IPKRPPLKLTOPPPPRN 346
DB 1934 SGKHNIENPPIVARYIRLHPTHYSIRSLRMELMGCDLNSMPLGMONKAISDSQITA 1993
QY 347 SNDLKN--TTAPPKIAKRAPKFTQPLQPRSSNEFFAQTEQTASPDIRNTVTVPNT 402
DB 1994 SSHLSNIFATWSPSOARLHLQGRTNARPRVS-----SABEMLQVDLQKTVKVTGIT 2045
RESULT 12
ID NRP2_RAT STANDARD; PRT; 925 AA.
AC Q35276;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
GN NRP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=97433085; PubMed=9288754;
RA Kolodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J., Ginty D.D.;
RT "Neuropilin is a semaphorin III receptor.";
RL Cell 90:753-762(1997).
CC -!- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165 AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: FOUND IN CERTAIN NEURONAL POPULATIONS OF THE CNS AND IN OTHER NONNEURONAL TISSUES INCLUDING MESENCHYMAL TISSUE LINING IN THE RIBS.
CC -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AF016297; AAC53338.1; --
 DR HSP; P12259; IC2T.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000421; FA58-C.
 DR InterPro; IPR000998; MAM_domain.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00629; MAM; 1.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00231; FA58C; 2.
 DR SMART; PS00600; MAM; 2; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01285; FA58C.1; 2.
 DR PROSITE; PS01286; FA58C.2; 2.
 DR Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor.
 KW SIGNAL 1 22
 FT CHAIN 23 925
 FT DOMAIN 23 858
 FT TRANSMEM 859 883
 FT DOMAIN 884 925
 FT DOMAIN 884 925
 FT DOMAIN 149 267
 FT DOMAIN 277 427
 FT DOMAIN 434 592
 FT DOMAIN 642 802
 FT DISULFID 28 55
 FT DISULFID 83 105
 FT DISULFID 149 175
 FT DISULFID 208 230
 FT DISULFID 277 427
 FT DISULFID 434 592
 FT CARBOHYD 152 152
 FT CARBOHYD 157 157
 FT CARBOHYD 629 629
 FT CARBOHYD 833 833
 FT CARBOHYD 834 834
 FT SEQUENCE 925 AA; 103896 MW; 38P62903F644851C CRC64;
 Query Match 8.5%; Score 295; DB 1; Length 925;
 Best Local Similarity 23.0%; Pred. No. 4.1e-13;
 Matches 134; Conservative 70; Mismatches 244; Indels 134; Gaps 22;
 QY 23 GKVCGLQWNN---HIESKGNIEITLPMGSHVSGRGLASVVDKQDLCLDTASN 79
 DB 102 GKHC-----NIAPPTIISGSVLYIKFTSDYARQAGSFLRVEIFKTS-----EDCSKN 152
 QY 80 FLEPFSKYCPAGCLLPFAEISGTIPGHVRSPLCMAGVHAGVSVNTLGGQIS----- 133
 DB 153 FTSFNGTIESPG-----FPE---KYPHNL-DCTFTILAKPRWEIIQLFTFDEHPLQV 203
 QY 134 -----VVISKGP-----YYESSLANNVTSVGHLSLSTLFTKTS----- 168
 DB 204 GEGDCKYDLDWDGIPHVGLPGKCGTKPTSKRSSTGILS---LTPHTDMAVAKDGF 260
 QY 169 -----CCYGLTHESGVADPDQITASSVLEWTHDTQENSWKPKARLKK 213
 DB 261 SARYLVHQPPEPNFQCNAPLGSRIANEQISASS-----TFSGRTWTPQOQSLHG 313
 QY 214 PGPPWAAFAFDEYQWLQIDLNKEKKTIGITTC--SPWVEHNYVVSAYRLTYSDDGQKW 271
 DB 314 DNGWTPNVDNSKEYLQVDLRFMTLTAIQAGISRETKGYVKSYLEVSTNGEDWM 373
 QY 272 YRPGVEQDKIFQGNKDYQDVRNNEFLPIIARFVNPTQOQKIAMKMLLGQFIP 331
 DB 374 YRHG--KNHKVQANNDATELVNLKLTPLTRFIRPQTHLGLALRLEFGC----- 427
 QY 332 KGRPKLTQPPPRNSNDLNTTAPPKIAGRAPK-----TQPLQPRSSNEFFPAQTEQ 386
 DB 428 -----RVTDAPCSNNLGMGLIADTQISASSTREYLVSPSARLVSSRSRGWFPNPAQ 482
 QY 387 TASPDIRNTVTNVTQDV-----ALAAVLVPLVNVLTLLILVCAWHW----- 432

DB 483 PGEENLQVDLGTPTKVKYGLIQARGGDSITAMEARAFVRKFKVSYSLNGKDEYIQDPR 542
 QY 433 RNRKKTEGT--YDLPYWDRAWMKQKPLPAKAVDHEETPVRYSSSESVNHLSPREVTT 490
 DB 543 TQQPKLFEGNMHYDTP-----DIRRF-----EPVPAQYRVYPERNSPAGIGM 585
 QY 491 VLQADSAEY--AQLYGVGIVGTLHQRSTFKP-----BEGKEAG 526
 DB 586 RLEVLGCDWTDTSKPTVETLGTPTVKSEETTPYPMDEDATECG 627
 RESULT 13
 NRP2_HUMAN
 ID NRP2_HUMAN STANDARD; PRT; 931 AA.
 AC 060462; O14820; O14821;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
 GN NRP2 OR VEGF165R2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
 RX MEDLINE=97470888; PubMed=9331348;
 RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
 RT "Neuropilin-2, a novel member of the neuropilin family, is a high affinity receptor for the semaphorins Sema E and Sema IV but not Sema III.";
 RT Neuron 19:547-559(1997).
 RL [2]
 RN SEQUENCE FROM N.A. (ISOFORM A22).
 RP TISSUE=Breast;
 RX MEDLINE=98188099; PubMed=9529250;
 RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
 RT "Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor.";
 RT Cell 92:735-745(1998).
 RL [3]
 RN CHARACTERIZATION
 RP MEDLINE=20309748; PubMed=10748121;
 RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
 RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid form of vascular endothelial growth factor (VEGF) and of placenta growth factor-2, but only neuropilin-2 functions as a receptor for the 145-amino acid form of VEGF.";
 RT J. Biol. Chem. 275:18040-18045(2000).
 RL [1] FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165 AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
 CC [1] SUBUNIT: NEUROFILIN-2 PROBABLY FORMS A HETEROMERIC COMPLEX WITH NEUROFILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN 3C RECEPTOR.
 CC [1] SUBCELLULAR LOCATION: Type I membrane protein.
 CC [1] ALTERNATIVE PRODUCTS: 3 ISOFORMS: A22 (SHOWN HERE), A0 AND A17;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC [1] SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
 CC [1] SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC [1] SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC [1] SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF022859; AAC51788.1; --
 DR EMBL; AF022860; AAC51789.1; --
 DR EMBL; AF016098; AAC12922.1; --

```
DR HSP; P12259; 1CZT.
DR Genes: HGNC:8005; NRP2.
DR MLM: 602070;
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR000998; FAM_domain.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00231; FA58C; 2.
DR SMART: SM00137; MAM; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
DR PROSITE: PS00600; MAM_2; 1.
KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Alternative splicing.
FT SIGNAL 1 20 OR 22 (POTENTIAL).
FT CHAIN 21 931 NEUROFILIN-2.
FT DOMAIN 21 864 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 865 889 POTENTIAL.
FT DOMAIN 890 931 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 28 142 CUB 1.
FT DOMAIN 149 267 CUB 2.
FT DOMAIN 277 427 F5/F8 TYPE C 1.
FT DOMAIN 434 592 F5/F8 TYPE C 2.
FT DOMAIN 642 802 MAM.
FT DOMAIN 671 674 POLY-SER.
FT DISULFID 28 55 BY SIMILARITY.
FT DISULFID 83 105 BY SIMILARITY.
FT DISULFID 149 175 BY SIMILARITY.
FT DISULFID 208 230 BY SIMILARITY.
FT DISULFID 277 427 BY SIMILARITY.
FT DISULFID 434 592 BY SIMILARITY.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 809 813 MISSING (IN ISOFORM A17).
FT VARSPIC 809 830 MISSING (IN ISOFORM A0).
FT CONFLICT 602 602 E -> K (IN REF. 1).
SQ SEQUENCE 931 AA; 104830 MW; 270CBAE69A0A797C CRC64;

Query Match 8.5%; Score 295; DB 1; Length 931;
Best Local Similarity 23.08; Pred. No. 4.2e-13;
Matches 133; Conservative 72; Mismatches 216; Indels 128; Gaps 21;

QY 23 GKCYGLGLOMN---HSIESKGNETILFLFMGIRVSGRGLASYSVIDKDLITCLDTASN 79
D 102 GKHC-----NIAPPTIISGSMYIKFTSDYARQAGAGFLRYEFTKGS-----EDCSKN 152
QY 80 FLEPEFSKYCPAGCLLPFAEISGTPHGYRDSPLCMHAGVAGVSVNLTGGQIS----- 133
D 153 FTSNPGTIESPG-----FPE---KYPHNL-DCTFTILAKPMETILQFLIFDLEHDLQV 203
QY 134 -----VVISKIP-----YVESLANNVTSVVGHLSTSLF----- 164
D 204 GEGDCKYDMLDWDGIPHVGLICKYCTKTPSELRSSTGLSTLSTFFHDMVAKDGFSA 263
QY 165 -----FKTSGCYGLGMSGVADPQITASSVLEWTDHTGQENSWKPKARLKPKGP 216
D 264 YLVHQEPLNFQCNVPLGMESGRINEQISASS-----TYSDGRTWPQOSRLHGDDN 316
QY 217 PWAFAFDEQWLQIDLNKKEKTIITG---STWVEHNYTVSYAIRLYSDGQKWTVYR 274
D 317 GWTPLNDSNKEYLQVLDLFTLMTAIAITQGAISRETQNGYVYKSVLEVTNGEDWMYR 376
QY 275 EPGVEODKIFGQNKDYHODVRRNNFLPLIARFIRVNPQWQOKIAMKMLLGCQFIPAGR 334
D 377 HG---KNHKVFOANDATEVVLNUKHAHLLTRFVIRPOTWSHGIALRLLELFGC----- 427
QY 335 PPKLTOPPPRNSDLKNTTAPPKIAKRAPKF-----TOPLQPRSSNEFFAQTEQTAS 389
```

```
Db 428 --RVTDAFCSNMLGMLGLIADQSISASTOYELWSPSAARLVSSRGWFFPPIQAPQCE 485
QY 390 PDIRNTVTPNVTQKV-----ALAAVLVPVLVMTLILILVCAHW-----RNR 435
D 486 EWLQVDLGTPTKTVKGVIIQGGAGGDSITAVEARAFVKVSYSLNGKDWYIQDPRTQQ 545
QY 436 KKKTEGT--YDLPYWDRAKWKMGKQFLPAKAVDHEETPVRYSSSEVNHLSPREYTTVLO 493
D 546 PKLFGNNHYDTP-----DIRRDP-----IPAQYRVVYPERWSPAGIMRL 588
QY 494 ADSAEY--AQLVGGIVGTLHORSTFKP-----EEGKEAG 526
D 599 VLGCMDTDSKPTVETLGTPTKSEETTPPYTEATECG 627

RESULT 14
NRP2_MOUSE
ID NRP2_MOUSE STANDARD; PRT: 931 AA.
AC O35375; O35373; O35374; O35376; O35377; O35378;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165 receptor 2).
GN NRP2
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX STRAIN=BALB/C;
MEDLINE=97470888; PubMed=9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III.";
RL Neuron 19:547-559(1997).
CC -!- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC -!- SUBUNIT: NEUROFILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC NEUROFILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN E RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: 6 ISOFORMS: A22 (SHOWN HERE), A0, A5, A17,
CC A0 AND B5; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING CNS, PNS AND IN SOME
CC NONNEURAL TISSUES INCLUDING LIMB BUDS, DEVELOPING BONES, MUSCLES,
CC INTESTINAL EPITHELIUM, KIDNEY, LUNG AND SUBMANDIBULAR GLAND.
CC -!- DEVELOPMENTAL STAGE: THE EXPRESSION PATTERN IS VERY DYNAMIC AND
CC IS DEVELOPMENTALLY REGULATED.
CC -!- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF022856; AAC53379.1; -
CC EMBL; AF022854; AAC53377.1; -
CC EMBL; AF022855; AAC53378.1; -
CC EMBL; AF022857; AAC53380.1; -
CC EMBL; AF022858; AAC53381.1; -
CC EMBL; AF022861; AAC53382.1; -
CC HSSP; P12259; 1CZT.
CC MGI; MGI:1100492; Nrp2.
CC InterPro; IPR000859; CUB_domain.
```


DR HSP; P12259; ICZT.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR000998; MAM_domain.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS00660; MAM_2; 1.
KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Cell adhesion.
FT SIGNAL 1 18
FT CHAIN 19 914
FT DOMAIN 20 847
FT TRANSFEM 848 870
FT DOMAIN 871 914
FT DOMAIN 25 139
FT DOMAIN 145 263
FT DOMAIN 273 422
FT DOMAIN 429 581
FT DOMAIN 636 801
FT DISULFID 25 52
FT DISULFID 80 102
FT DISULFID 145 171
FT DISULFID 204 226
FT DISULFID 273 422
FT DISULFID 429 581
SQ SEQUENCE 914 AA; 102480 MW; DD2EE6D6F0CB68C CRC64;

Query Match 8.2%; Score 284.5; DB 1; Length 914;
Best Local Similarity 21.7%; Pred. No. 2.3e-12;
Matches 135; Conservative 79; Mismatches 190; Indels 219; Gaps 24;
QY 17 DAGAQQKYGKGLGLOMNH-IESKGNELTLFMSGIHVSGRGFLASYVIDKDLITCLD 75
DB 217 DVGPHIGRYCG--QNNPGRVRSIGILSMVFYTDISAIAKEGFSANTSV----- 262
QY 76 TASFLEPEFSKYCPAGCLLPFAISGITHGYRDSPLCMAGVHAGVVSNTLG----- 129
DB 263 -SQSSVSEDFQ-----CMEPLGMEGEI-----HSDQITVSSQISAIWSSERSRLNYPE 310
QY 130 -----GOISV-----VISKGI--PYESSLANNVTS----- 153
DB 311 NGWTPGEDSVREWIQVDLGLLRFVSGIGTQCAISKETKKEYLKYRYVDVSSNGEDWITL 370
QY 154 -----VVGHLSTSLFTFKTSG-----C 170
DB 371 KEGNKPVVFGNSNPTDVVYRPFKPVLTREVRIPKPSWENGVSLL-RPEVYCKITDYP 429
QY 171 YGTLMESGVIAADQITASSVLEWTDHTGOENSKPKKARKLKPGPWAAATDE---YQ 227
DB 430 SCMLQWVSLIPDSQITAS-----TQDRNWIPEARLITSRSGWALPPTHPTNE 481
QY 228 WLQIDLANKKKITGITGTSTWVHNHYVSAYRILYSDGOKWTYVREPGEVDQKIFQGN 287
DB 482 WLQIDLGEKIVRGIIVGGKHKRENKVFMKFKIGYSNNGSDWKIMDSSKKKIKTFEGN 541
QY 288 KDYHQDVNNFLPLIARFIRVNPQ-WQKIAMKMLLGCQFIPKGRPPKLTQPPPPRN 346
DB 542 TNYDTPELRTF-EPVSTRIIRVYPERATHAGLGLRMELLGC----- 581
QY 347 SNDLKNTTAPPKIAKGRAPKTPQLPPRSSNEFFAQTEQTTASPDIRNTVTTPNWKDVA 406
DB 582 --ELEAPTAFTVSEGPVDECDQDQANCSHGSGTGTTLNTEKPTVIDNTVOPE----- 633
QY 407 LAAVLPVPLVMVLATLILILVCAMHWRNRKKTGTGVDL-----PYWDRAG 452

Db 634 ----LPP-----YNLNCGFGGSHKTLQWEHDNQVDLKWAILTSKGTGPIODHT- 678
QY 453 WVKGMKQFLPAKAVDHEE-----TPVRYSSSEVN-----HLSPREVTTVLOADSAEY 499
DB 679 ---GDGNFIYSQADESQKGVARLLSPIIYSSQNSAHCMFTFWHMS----- 720
QY 500 AOLVGGIVGTLHORSTF-KPEE 521
DB 721 -----GPHVGTILKIKLYQRPDE 738

Search completed: January 21, 2003, 09:55:40
Job time : 16.0326 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:42:00 ; Search time 25.0427 Seconds
(without alignments)
5372.776 Million cell updates/sec

Title: US-10-060-830-1114
Perfect score: 3470
Sequence: 1 MFLFLLLVLLVLLLEDAGA.....TQEVSGAGRGDCDVFKEIL 653

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- SPTREMBL_21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertibrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3422	98.6	775	4 Q96PD2	Q96pd2 homo sapien
2	3422	98.6	775	4 Q8TDX2	Q8tdx2 homo sapien
3	2984.5	86.0	769	11 Q91ZV2	Q91zv2 ratus norv
4	2950.5	85.0	769	11 Q91ZV3	Q91zv3 mus musculus
5	1151.5	33.2	251	11 Q9D9K5	Q9d9k5 mus musculus
6	984.5	28.4	364	4 Q14089	Q14089 homo sapien
7	417	12.0	460	11 Q9D696	Q9d696 mus musculus
8	414	11.9	432	11 Q9R327	Q9r327 mus musculus
9	411	11.8	503	11 Q9D4J3	Q9d4j3 mus musculus
10	359	10.3	2119	13 Q90X47	Q90x47 brachydanio
11	355.5	10.2	2183	11 Q88783	Q88783 mus musculus
12	355	10.2	480	4 Q43854	Q43854 homo sapien
13	352	10.1	480	11 Q35474	Q35474 mus musculus
14	331	10.1	2224	4 Q43737	Q43737 homo sapien
15	323.5	9.3	426	11 Q9WTS3	Q9wts3 mus musculus
16	321	9.3	463	11 Q9R1X9	Q9r1x9 mus musculus

17	317	9.1	2343	6 O18806	O18806 canis famil
18	311	9.0	2343	6 O62730	O62730 canis famil
19	308.5	8.9	216	4 Q14286	Q14286 homo sapien
20	295	8.5	901	4 Q9H2E4	Q9h2e4 homo sapien
21	295	8.5	901	4 Q9H2D5	Q9h2d5 homo sapien
22	295	8.5	906	4 Q9H2E3	Q9h2e3 homo sapien
23	295	8.5	906	4 Q9H2D4	Q9h2d4 homo sapien
24	294	8.5	926	11 Q8QZY7	Q8qzy7 mus musculu
25	288.5	8.3	919	13 Q8UVRO	Q8uvr0 gallus gall
26	288.5	8.3	936	13 Q8UVQ9	Q8uvq9 gallus gall
27	286.5	8.3	923	13 Q8QFX6	Q8qfx6 brachydanio
28	285.5	8.2	555	4 Q9H2E2	Q9h2e2 homo sapien
29	271	7.8	363	6 Q77718	Q77718 equus cabal
30	271	7.8	921	11 Q9QX38	Q9qx38 ratus norv
31	269.5	7.8	224	11 Q9R1M6	Q9r1m6 mus musculu
32	267.5	7.7	858	5 Q76470	Q76470 lytechinus
33	266.5	7.7	609	4 Q96190	Q96190 homo sapien
34	265	7.6	644	4 Q961H5	Q961h5 homo sapien
35	265	7.6	704	4 Q9H2E1	Q9h2e1 homo sapien
36	265	7.6	779	4 Q9BTL9	Q9btl9 homo sapien
37	228	6.6	335	4 Q9BTL9	Q9btl9 homo sapien
38	227.5	6.6	1128	11 Q88442	Q88442 mus musculu
39	224.5	6.5	1158	4 Q14113	Q14113 homo sapien
40	220.5	6.4	764	11 Q9D2L5	Q9d2l5 mus musculu
41	220.5	6.4	764	11 Q54860	Q54860 mus musculu
42	220.5	6.4	764	11 Q8VDQ4	Q8vdq4 mus musculu
43	208	6.0	728	6 Q97567	Q97567 bos taurus
44	204	5.9	719	11 Q61281	Q61281 mus musculu
45	204	5.9	721	4 Q96M80	Q96m80 homo sapien

ALIGNMENTS

RESULT 1

Q96PD2 PRELIMINARY; PRT; 775 AA.
AC Q96PD2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like protein.
DE ESDN.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11447234;
RA Matsumori A., Sasayama S., Honjo T., Tashiro K., Ohashi N., Kobuke K., Furukawa Y., Sugai M., Tanigaki K.,
RT "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is Up-regulated after Vascular Injury.",
RT J. Biol. Chem. 276:34105-34114(2001).
RL -!- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL REGULATORS.
CC EMBL; AF387547; AAL30178.1; -!
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1
DR Pfam; PF00754; F5_F8_type_C; 1.
DR PROSITE; PS01180; CUB; 1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 775 AA; 85034 MW; 3D06F81EF2337010 CRC64;

Query Match 98.6%; Score 3422; DB 4; Length 775;
Best Local Similarity 89.6%; Pred. No. 1.6e-273;
Matches 653; Conservative 0; Mismatches 0; Indels 76; Gaps 1;

```

QY 1 MPEFLLLVLLVLLLEDAGAAQ----- 22
Db 47 MPEFLLLVLLVLLLEDAGAAQGGCGHTVLPESGTLTINYPQTPNSTVCEWEIRVK 106
QY 23 -----GKYCGLGLQMNHSIESKGNIT 44
Db 107 MGERVRIKFGDFIEDSDSCHFNRYLRIYNGIVGSRTETGKYCGLGLQMNHSIESKGNIT 166
QY 45 LLEMSGIHVSGRGLASYSVIDKQDLTCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 104
Db 167 LLEMSGIHVSGRGLASYSVIDKQDLTCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 226
QY 105 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFT 164
Db 227 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFT 286
QY 165 FKTSGCYGTGLMESGVITADPOITASSVLEWTDHTGQENSWKPKKARLKKPGPPWAAAFATD 224
Db 287 FKTSGCYGTGLMESGVITADPOITASSVLEWTDHTGQENSWKPKKARLKKPGPPWAAAFATD 346
QY 345 RNSNDLKNNTAPPKIAKGRAPKFTQPLQPRSSNEFFPAQTEOTTASPDIRNTVTPNVTKD 404
Db 467 RNSNDLKNNTAPPKIAKGRAPKFTQPLQPRSSNEFFPAQTEOTTASPDIRNTVTPNVTKD 526
QY 405 VALAAVLVPLVWVLTLLILVCAWHNRNKKKTEGYDLPYWDRAWGKMGKQFLPAK 464
Db 527 VALAAVLVPLVWVLTLLILVCAWHNRNKKKTEGYDLPYWDRAWGKMGKQFLPAK 586
QY 465 AVDHEETPVRYSSSEVNHLSPREVTTLQADSAEYAQPLVGGIVGTLHORSTFFKPEEGKE 524
Db 587 AVDHEETPVRYSSSEVNHLSPREVTTLQADSAEYAQPLVGGIVGTLHORSTFFKPEEGKE 646
QY 525 AGYADLPYNSPGQEVYHAYAEPLIPITGPEYATPIIMDSGHPTTSVGQPSSTFFKATGN 584
Db 707 QPPPLVGTNTLLSRTDSCSSAAQAYDTPKAGKPGLPAPDELVYQVPOSTQEVSGAGRDG 766
QY 645 ECDVFKEL 653
Db 767 ECDVFKEL 775

RESULT 2
Q8TDX2 PRELIMINARY; PRT; 775 AA.
ID Q8TDX2 AC Q8TDX2;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CLCP1.
GN CLCP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN NCBI_TaxID=9606;
RP [1]
RC SEQUENCE FROM N.A.
RA Koshikawa K., Osada H., Kozaki K., Konishi H., Masuda A.,
RA Tateatsu Y., Mitsudomi T., Nakao A., Takahashi T.;
RT *Significant up-regulation of a novel gene, CLCP1, in a highly
RT metastatic lung cancer subline as well as in lung cancers in vivo.*;

```

```

RL Oncogene 0:0-0(2002).
DR EMBL: AB073146; BAB91138.1; --; 6D07223B21BE5A42 CRC64;
SQ SEQUENCE 775 AA; 85073 MW; 6D07223B21BE5A42 CRC64;

Query Match 98.6%; Score 3422; DB 4; Length 775;
Best Local Similarity 89.6%; Pred. No. 1.6e-273;
Matches 653; Conservative 0; Mismatches 0; Indels 76; Gaps 1;

QY 1 MPEFLLLVLLVLLLEDAGAAQ----- 22
Db 47 MPEFLLLVLLVLLLEDAGAAQGGCGHTVLPESGTLTINYPQTPNSTVCEWEIRVK 106
QY 23 -----GKYCGLGLQMNHSIESKGNIT 44
Db 107 MGERVRIKFGDFIEDSDSCHFNRYLRIYNGIVGSRTETGKYCGLGLQMNHSIESKGNIT 166
QY 45 LLEMSGIHVSGRGLASYSVIDKQDLTCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 104
Db 167 LLEMSGIHVSGRGLASYSVIDKQDLTCLDTASNFLEPEFSKYCPAGCLLPFAEISGTI 226
QY 105 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFT 164
Db 227 PHGYRDSPLCMAGVHAGVSVNTLGGQISVVISKGIPIYESSLANNVTSVVGHLSTSLFT 286
QY 165 FKTSGCYGTGLMESGVITADPOITASSVLEWTDHTGQENSWKPKKARLKKPGPPWAAAFATD 224
Db 287 FKTSGCYGTGLMESGVITADPOITASSVLEWTDHTGQENSWKPKKARLKKPGPPWAAAFATD 346
QY 345 RNSNDLKNNTAPPKIAKGRAPKFTQPLQPRSSNEFFPAQTEOTTASPDIRNTVTPNVTKD 404
Db 467 RNSNDLKNNTAPPKIAKGRAPKFTQPLQPRSSNEFFPAQTEOTTASPDIRNTVTPNVTKD 526
QY 405 VALAAVLVPLVWVLTLLILVCAWHNRNKKKTEGYDLPYWDRAWGKMGKQFLPAK 464
Db 527 VALAAVLVPLVWVLTLLILVCAWHNRNKKKTEGYDLPYWDRAWGKMGKQFLPAK 586
QY 465 AVDHEETPVRYSSSEVNHLSPREVTTLQADSAEYAQPLVGGIVGTLHORSTFFKPEEGKE 524
Db 587 AVDHEETPVRYSSSEVNHLSPREVTTLQADSAEYAQPLVGGIVGTLHORSTFFKPEEGKE 646
QY 525 AGYADLPYNSPGQEVYHAYAEPLIPITGPEYATPIIMDSGHPTTSVGQPSSTFFKATGN 584
Db 647 AGYADLPYNSPGQEVYHAYAEPLIPITGPEYATPIIMDSGHPTTSVGQPSSTFFKATGN 706
QY 585 OPPPLVGTNTLLSRTDSCSSAAQAYDTPKAGKPGLPAPDELVYQVPOSTQEVSGAGRDG 644
Db 707 QPPPLVGTNTLLSRTDSCSSAAQAYDTPKAGKPGLPAPDELVYQVPOSTQEVSGAGRDG 766
QY 645 ECDVFKEL 653
Db 767 ECDVFKEL 775

RESULT 3
Q912V2 PRELIMINARY; PRT; 769 AA.
ID Q912V2 AC Q912V2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like
DE protein.
GN ESDN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAMLEY;
RX PubMed=11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT *ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury."
RL J. Biol. Chem. 276:34105-34114(2001).
CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AF387549; AAL30180.1; -.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR004043; LCCL_dom.
DR Pfam: PF00431; CUB; 1.
DR Pfam: PF00754; F5_F8_type_C; 1.
DR PROSITE: PS01180; CUB; 1.
DR DNA-binding; Transcription regulation.
KW SEQUENCE 769 AA; 83915 MW; 5B7936C8FA063A93 CRC64;

Query Match 86.0%; Score 2984.5; DB 11; Length 769;
Best Local Similarity 78.3%; Pred. No. 28-237;
Matches 566; Conservative 36; Mismatches 42; Indels 79; Gaps 3;

QY 7 LLLVLLLLLEDAGAQQ-----GKYGGLGLQNMHSIESKGNITLLFMSG 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 50 LLLLLLLLPDGAQKGDGCGHTVLPESGTLTSINYPHTYVCKWEIRVKTGERIR 109
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 23 -----GKYGGLGLQNMHSIESKGNITLLFMSG 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 110 IKFGDFIEDSDYCHLNLYKIFNGISVRSYRTEIGYKGLQNMHSIESKGNITLLFMSG 169
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 51 IHVSGRGFLASVYIDKDLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTHPGYRD 110
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 170 IHASGRGFLASVYIDKDLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTHPGYRD 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 111 SSPLCMAGHAGVSVNTLGGQISVTSKIPYESSLANNVTSVGHLSLSTLFTKTSGC 170
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 230 SSPLCMAGHAGVSVNTLGGQISVTSKIPYESSLANNVTSVGHLSLSTLFTKTSGC 289
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 171 YGLGMESGVADPQITASSVLEWTDHTGQENSWKPKARLKKPGPPWAAFAFDEYQWLQ 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 290 YGLGMESGVADPQITASSVLEWTDHTGQENSWKPKARLKKPGPPWAAFAFDEYQWLQ 349
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 231 IDLNKEKKITGITGTSWVHEHNYYSAYRILYSDGQKWTYVREPGVQDQKIFQGNKDY 290
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 350 IDLNKEKKITGITGTSWVHEHNYYSAYRILYSDGQKWTYVREPGVQDQKIFQGNKDY 409
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 291 HODVRNNEFLPIIARIRVNPQOQKIAMKELLGCOFTPKGRPKLTQPPPPRNSNDL 350
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 410 HKDVRNNEFLPIIARIRVNPQOQKIAMKELLGCOFTPKGRPKLTQPPPPRNSNDL 469
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 351 KNTAPPKIAKGRAPKFTQPLQPRSSNEFFPAQTEQTASPDIRNTTNTVNTKDVAAAV 410
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 470 KNTTVHPKL--GRAPKFTQALQPRSRNDLPLPAQTATPDVKNVTVTSPVKDVAAAV 527
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 411 LVPVLVAVLTLLILVCAWHNRNKKTEGTYDLPYWDKAGWKGKQFLPAKAYDHEE 470
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 528 LVPVLVAVLTLLILVCAWHNRNKKTEGTYDLPYWDKAGWKGKQFLPAKAYDHEE 587
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 471 TPVRSSESVNHLSPREVTTLQADSAYEAQPLVGGVIGTLHORSTFKPEEGKEAGYADL 530
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 588 TPVRSSESVNHLSPREVTTLQADSAYEAQPLVGGVIGTLHORSTFKPEEGKEAGYADL 647
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 531 DPYNSPGQEVYHAYAPLPTGPETATPIIMDSMGHPTTSVGOPSTSTFKATGNQPPPLV 590
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 648 DPYNSPGQEVYHAYAPLPTGPETATPIIMDSMGHPTTSVGOPSTSTFKATGNQPPPLV 707
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 591 GTYNTLLSRDSCSSAQYDTPKAGKPGLPAPDELVIQVPSQTSQVSGAGRDGCECVFK 650
```

```
Db 708 GTYNTLLSRDSCSSAQYDTPKGGKPA-AAPELVYQVPQSTQEASGAGRDEKFAFK 766
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 651 EIL 653
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 767 EYL 769
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
Q912V3 PRELIMINARY; PRT; 769 AA.
ID Q912V3
AC Q912V3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Endothelial and smooth muscle cell-derived neuropilin-like
DE protein.
GN ESDN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR;
RX PubMed=11447234;
RA Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT *ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT Up-regulated after Vascular Injury."
RL J. Biol. Chem. 276:34105-34114(2001).
CC -1- SIMILARITY: BELONGS TO THE CRP/FNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
DR EMBL; AF387548; AAL30179.1; -.
DR MGD; MGI:1920629; ESDN.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR004043; LCCL_dom.
DR Pfam: PF00431; CUB; 1.
DR Pfam: PF00754; F5_F8_type_C; 1.
DR PROSITE: PS01180; CUB; 1.
DR DNA-binding; Transcription regulation.
KW SEQUENCE 769 AA; 83774 MW; 73C1F1646FA3F017 CRC64;

Query Match 85.0%; Score 2950.5; DB 11; Length 769;
Best Local Similarity 77.3%; Pred. No. 1.3e-234;
Matches 559; Conservative 37; Mismatches 48; Indels 79; Gaps 3;

QY 7 LLLVLLLLLEDAGAQQ-----GKYGGLGLQNMHSIESKGNITLLFMSG 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 50 LLLLLLLLPDGAQKGDGCGHTVLPESGTLTSINYPHTYVCKWEIRVKTGERIR 109
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 23 -----GKYGGLGLQNMHSIESKGNITLLFMSG 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 110 IKFGDFIEDSDYCHLNLYKIFNGISVRSYRTEIGYKGLQNMHSIESKGNITLLFMSG 169
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 51 IHVSGRGFLASVYIDKDLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTHPGYRD 110
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 170 THAAGRGFLASVYIDKDLITCLDTASNFLEPEFSKYCPAGCLLPFAEISGTHPGYRD 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 111 SSPLCMAGHAGVSVNTLGGQISVTSKIPYESSLANNVTSVGHLSLSTLFTKTSGC 170
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 230 SSPLCMAGHAGVSVNTLGGQISVTSKIPYESSLANNVTSVGHLSLSTLFTKTSGC 289
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 171 YGLGMESGVADPQITASSVLEWTDHTGQENSWKPKARLKKPGPPWAAFAFDEYQWLQ 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 290 YGLGMESGVADPQITASSVLEWTDHTGQENSWKPKARLKKPGPPWAAFAFDEYQWLQ 349
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 231 IDLNKEKKITGITGTSWVHEHNYYSAYRILYSDGQKWTYVREPGVQDQKIFQGNKDY 290
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 350 IDLNKEKKITGITGTSWVHEHNYYSAYRILYSDGQKWTYVREPGVQDQKIFQGNKDY 409
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 291 HQDVRRNFPPIIARFTRVNPOTQOQIAKMKMELLGCOFTPKGRPPKLTQPPPPRNSNDL 350
DB 410 HKDVRNFPPIIARFTRVNPQVQOQIAKMKVELLGCQFTLKGRPLKT--PPRNGNLL 467
QY 351 KNTAPKIAKGRAPKFTQPLQPRSSNEEPAQTEOTTASDPDIRNTTVPNTKDVAAAV 410
DB 468 RNTARPKLKGKRAPKFTQVLPQPSRNLPLVPAETTTTIDIKNTTTPSTVKDVALAAV 527
QY 411 LVPVLVAVLTTLLILVCAWHNRNKKTKTGYDLPYWDRAAGWKMFKQFPAKAVDHEE 470
DB 528 LVPVLVAVLTTLLILVCAWHNRNKKTKTGYDLPYWDRAAGWKMFKQFPAKAVDHEE 587
QY 471 TPVRSSEVNHLSPREVTVLQADSAYEAQPLVGGIVGTLHQRSTFKPBEKGAGYADL 530
DB 588 TPVRYSTSEVSHLSAREVTVLQADSAYEAQPLVGGIVGTLHQRSTFKPBEKGAGYADL 647
QY 531 DPVNSPQOEVYHAYAEPLPTGPEYATPIIMDSCHPTTSVGQPSSTFKATGNQPPPLV 590
DB 648 DPVNSPQOEVYHAYAEPLPTGPEYATPIIMDSCHPTTSVGQPSSTFKATGNQPPPLV 707
QY 591 GTYNTLLSRTDSCSSAQAYDTPKAGKPGLPAPDELVYQVQPSQTEVSGAGDGCDFVK 650
DB 708 GTYNTLLSRTDSCSSAQAYDTPKAGKPGLPAPDELVYQVQPSQTEVSGAGDGCDFVK 766
QY 651 EIL 653
DB 767 EIL 769

RESULT 5
Q9D9K5 PRELIMINARY; PRT; 251 AA.
AC Q9D9K5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 170055P2IRik protein.
GS ESDN OR 1700055P2IRIK.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniwa M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK006805; BAB24750.1; -.
DR MGD; MGI:1920629; ESDN.
SQ SEQUENCE 251 AA; 27428 MW; F86E0AA15EF51AE9 CRC64;

Query Match 33.2%; Score 1151.5; DB 11; Length 251;
Best Local Similarity 87.6%; Pred. No. 7.6e-87;
Matches 218; Conservative 9; Mismatches 21; Indels 1; Gaps 1;
```

```
QY 405 VALAAVLVPVLVAVLTTLLILVCAWHNRNKKTKTGYDLPYWDRAAGWKMFKQFPAK 464
DB 4 VALAAVLVPVLVAVLTTLLILVCAWHNRNKKTKTGYDLPYWDRAAGWKMFKQFPAK 63
QY 465 AVOHEETPVRYSSSEVNHLSPREVTVLQADSAYEAQPLVGGIVGTLHQRSTFKPBEKKE 524
DB 64 SVDHEETPVRYSTSEVSHLSAREVTVLQADSAYEAQPLVGGIVGTLHQRSTFKPBEKKE 123
QY 525 AGVADLPYNSPQOEVYHAYAEPLPTGPEYATPIIMDSCHPTTSVGQPSSTFKATGN 584
DB 124 AGVADLPYNSPQOEVYHAYAEPLPTGPEYATPIIMDSCHPTTSVGQPSSTFKATGN 183
QY 585 QPPLVGTNTLLSRTDSCSSAQAYDTPKAGKPGLPAPDELVYQVQPSQTEVSGAGRDG 644
DB 184 QPHALVGTNTLLSRTDSCSSAQAYDTPKAGKPGLPAPDELVYQVQPSQTEVSGAGRDG 242
QY 645 ECDVFEKEIL 653
DB 243 KFDAFEKEIL 251

RESULT 6
Q14089 PRELIMINARY; PRT; 364 AA.
AC Q14089;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.0 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shibata T.;
RT "unpublished.";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL; D29810; BAA18909.1; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR004043; LCCL_dom.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR SMART; SM00042; CUB; 1.
DR PROSITE; PS01180; CUB; 1.
DR Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 364 AA; 39954 MW; FFCAG292B9E0801F CRC64;

Query Match 28.4%; Score 984.5; DB 4; Length 364;
Best Local Similarity 69.4%; Pred. No. 8.2e-73;
Matches 209; Conservative 16; Mismatches 47; Indels 29; Gaps 6;

QY 23 GKYGGLGLQMHNSIESKNEITLLFMSGIHVSGRGLASYSVIDKDLITCLDTASNEFL 82
DB 76 GKYGGLGLQMHNSIESKNEITLLFMSGIHVSGRGLASYSVIDKQ-----RSNYLF 127
QY 83 PEFSKY-----CPAGCLLPFAEISGTIPHGVRDSSPLCMAGVHAGVYVNTLGGQIS 133
DB 128 GRCIQFVNLSSVSTAPAGCLLPFAEISGTIPHGVRDSSPLCMAGVHAGVYVNTLGGQIS 187
QY 134 VVISKGIPIYESSLANNVTYVVGHLSTSL-PTFKTSQCYGTLGMESGVIAADPQTASSYL 192
DB 188 VVISKGIPIYESSLANNVTYVVGHLSTSL-PTFKTSQCYGTLGMESGVIAADPQTASSYL 247
QY 193 EWTDTGTGNSWPKPKARLKPFPAAAFATDEYQWLQIDLNKKEKITGIIITGSTMVHEH 252
DB 248 EWTDTGTGNSWPKPKARLKPFPAAAFATDEYQWLQIDLNKKEKITGIIITGSTMVHEH 306
QY 253 NYVVSAYRILYSDD-----GOKWTYVREPGEVDQKIFQGNKD-YHODVRNFPPIIARF 306
```


QY 242 IITGSTMVHNYVYAYRILYSDGQKWTYREPGVEODKIFOGNKDYHODVRNRLPP 301
 Db 173 -----PGM----- 175
 QY 302 IARFIRVNPTMOOKIAMKMLLGCQFIPKGRPKLTQPPPPRNSDLKNTAPPKIAK 361
 Db 176 ----- 175
 QY 362 GRAPKFTQLOPRSSNEFFPAQTEQTASPDIRNTVTPNVTKDVALAALVPLVMVLT 421
 Db 176 -----NIT-TVAIPSVI--FIALLTG 194
 QY 422 LILILVCAWHNRKTKTEGYDLPY-----WDRAGWKGKQFLPAKAVDHEETP-VRYS 476
 Db 195 MGIFAIC-----RRKKKGN---PYVSADAQKTCGWKQIKY-----PFARHQ 233
 QY 477 SSE--VNHLSPREVT---TVLOADSAEYAOPLVGGIVGTLHORSTFKP-----EE---G 522
 Db 234 STEFTISYDNEKEMTKLDTSDMADYQQLMIGTGTVARKGSTFRPMDTDEVRVNT 293
 QY 523 KEAGYADLPYNSGQEVYHAYAEPLPTGPPEYATPII-----MDMSGHPTTSVG 572
 Db 294 EASGH-----YDCPHRPGHEYALPLTHSEPEYATPIVERHLLRAHTFSTQSGYRVPG-P 347
 QY 573 QPSTSTFKATGNQPPPLVGTNTLLSRDSCSSAQAOYDTPKAGKPLPAPDELVYQVPQ 632
 Db 348 RPTUKHSHSGGFFPATGATQIESYQRPASPKPVGGYDKP-AASSFLDSRD-----PA 400
 QY 633 STQEVSGAGRDG 644
 Db 401 SQSQMTSGGDDG 412

RESULT 9

Q9D4J3 ID Q9D4J3 PRELIMINARY; PRT; 503 AA.
 AC Q9D4J3
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE 4631413K11Rik protein.
 GN 4631413K11Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21083660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuhl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 FT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 CC EMBL; AK016485; BAB30265.1; -.

DR MGD; MGI:1913936; 4631413K11Rik.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR004043; LCC1_dom.
 DR Pfam; PF00431; CUB; 1.
 DR SMART; SM00042; CUB; 1.
 DR PROSITE; PS01180; CUB; 1.
 SQ SEQUENCE 503 AA; 54547 MW; FEB121E845CAA06B CRC64;

Query Match 11.8%; Score 411; DB 11; Length 503;

Best Local Similarity 22.2%; Pred. No. 3e-25;

Matches 147; Conservative 70; Mismatches 148; Indels 296; Gaps 19;

QY 13 LLEDAGAOQKCYCGICGLQNMNHSIESKNEITLLFMGSIHVSGRFLASYVIDKQDLIT 72
 Db 90 LIFSATDQYGPYCG-SWAPVKELRLNSNEVTLFSGSHISGRFLTYASSDPDLIT 148
 QY 73 CLDTASNFLEPFESKCYCPAGCLLPFAETISGTPHGYRDSPLCMAGVHGVVSTLGGQI 132
 Db 149 CLERGSYFEEKYSKFCPCAGCDTAGDISGNTKGYRDTSLCKAAIHAGIITDELGHI 208
 QY 133 SVTISKGIPIYESSLANNVTSVGHLSLSTLFTFTKSGCYGLGMESGVIAADPQTASSVL 192
 Db 209 NLLQSKGISHYEGLLANGVLNRHSGLSLSEKRLFTT----- 243
 QY 193 EWTDTGQENSWKPKARLKKPGPPWAFATDEYQWLQIDLNKKKKTGITTGSTWVEH 252
 Db 244 ----- 243
 QY 253 NYTVSAVILYSDGQKWTYREPGVEODKIFOGNKDYHODVRNRLPPIIARFIRNPT 312
 Db 244 -----PGM----- 246
 QY 313 QWQOKIAMKMLLGCQFIPKGRPKLTQPPPPRNSDLKNTAPPKIAKGRAPKFTQPLQ 372
 Db 247 ----- 246
 QY 373 PRSSNEFFPAQTEQTASPDIRNTVTPNVTKDVALAALVPLVMVLTLLILVCAWH 432
 Db 247 -----NIT-TVAIPSVI--FIALLTGMGIFAIC----- 272
 QY 433 RNKKKTKTEGYDLPY-----WDRAGWKGKQFLPAKAVDHEETP-VRYSSE--VNHLS 485
 Db 273 --RRKKKGN---PYVSADAQKTCGWKQIKY-----PFARHSTEEFTISYDNE 315
 QY 486 REVTV---TVLOADSAEYAOPLVGGIVGTLHORSTFKP-----EE---GKEAGYADLPY 533
 Db 316 KEMTKLDTSDMADYQQLMIGTGTVARKGSTFRPMDTDEVRVNTASGH-----Y 370
 QY 534 NSPGQEVYHAYAEPLPTGPPEYATPII-----MDMSGHPTTSVGQPSSTFFKATG 583
 Db 371 DCPHPRGHEYALPLTHSEPEYATPIVERHLLRAHTFSTQSGYRVPG-PRPTHKSHSG 429
 QY 584 NQPPPLVGTNTLLSRDSCSSAQAOYDTPKAGKPLPAPDELVYQVPQSTQEVSGAGRD 643
 Db 430 GFPPATGATQVESTQRPASPRPVGGYDKP-AASSFLDSRD-----PASQMTSGGDD 482
 QY 644 G 644
 Db 483 G 483

RESULT 10

Q90X47 ID Q90X47 PRELIMINARY; PRT; 2119 AA.
 AC Q90X47
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SC:B22015.3 (novel protein similar to vertebrate coagulation factor V
 DE DE and VIIII).
 GN SC:B22015.3.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Query Match	10.28;	Score 355.5;	DB 11;	Length 2183;
Best Local Similarity	40.18;	Pred. No. 1.1e-19;		
Matches	79;	Conservative 28;	Mismatches 75;	Indels 15;
				Gaps 4;

Qy	135	VISKGIPYESSIANVNTSVVGHLSLSTLFTFKYSCGYTLGHMGSVIADPQITASSVLE-	193
Db	1995	IVARYIRIHPTKSYNRP-----LRLELQGCYNGCSTPLGLDGRIDQKQITASSFKKS	2049
Qy	194	-WDHTCGENSWPKPKARKLKKPG--PPAAAFATDEYQWLOIDLNEKKITGITITGSTMV	250
Db	2050	WWDGY-----WPSLARLNAQGRVNAQKANNKQWLOVLLAKKVKTAIVTGCKSL	2103
Qy	251	EHNYYVSAYRILYSDGQKWTVYREPGVEQDKIFQGNKDYHDYVRNFPPIIARFIRVN	310
Db	2104	SSEMYVKSYIQSYDQGVAKMPYRKQSSMVDKIFEGNSNTKGMKNFFNPPIISIRFIRI	2163
Qy	311	PTQWQKQIAKMWELGCG 327	
Db	2164	PKTNQSIARLRLFLFC 2180	
RESULT	12		
ID	043854	PRELIMINARY; PRT; 480 AA.	
AC	043854; 043855;		
DT	01-JUN-1998 (T-EMBLrel. 06, Created)		
DT	01-JUN-1998 (T-EMBLrel. 06, Last sequence update)		
DT	01-DEC-2001 (T-EMBLrel. 19, Last annotation update)		
DE	Integrin-binding protein DELI precursor.		
GN	DELI.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_Taxid:9606;		
RP	[1]		
RN	SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.		
RC	TISSUE=EMBRYONIC LUNG;		
RX	MEDLINE=98083109; PubMed=9420328;		
RA	Hidai C., Zupancic T.J., Penta K., Mikhail A., Kawana M.,		
RA	Quertbach R., Hogan B.L.M., Snodgrass R., Quertbach T.,		
RT	"Cloning and characterization of developmental endothelial locus-1: an		
RT	embryonic endothelial cell protein that binds the alphavbeta3 integrin		
RT	receptor";		
RL	Genes Dev. 12:21-33(1998).		
CC	-1- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH		
CC	INTERACTION WITH THE ALPHA-V BETA3 INTEGRIN RECEPTOR. INHIBITS		
CC	FORMATION OF VASCULAR-LIKE STRUCTURES. MAY BE INVOLVED IN		
CC	REGULATION OF VASCULAR MORPHOGENESIS OF REMODELING IN EMBRYONIC		
CC	DEVELOPMENT		
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.		
CC	-1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: LONG ISOFORM (SHOWN HERE) AND		
CC	SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.		
CC	-1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.		
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.		
DR	EMBL; U70312; AAC02648.1; -		
DR	EMBL; U70313; AAC02649.1; -		
DR	HSSP; P12259; ICZT.		
DR	InterPro; IPR000152; Asx_hydroxyl.		
DR	InterPro; IPR000561; EGF-like.		
DR	InterPro; IPR000742; EGF_2		
DR	InterPro; IPR001881; EGF_Ca.		
DR	InterPro; IPR000421; FA58_C.		
DR	pfam; PF00008; EGF_3.		
DR	pfam; PF00754; F5_F8_type_C; 2.		
DR	SMART; SM00179; EGF_CA; 1.		
DR	SMART; SM00001; EGF-like; 2.		
DR	SMART; SM00231; FA58C; 2.		
DR	PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.		
DR	PROSITE; PS00022; EGF_1; UNKNOWN_2.		
DR	PROSITE; PS01186; EGF_2; UNKNOWN_2.		
DR	PROSITE; PS01187; EGF_CA; 1.		
DR	PROSITE; PS01285; FA58C_1; 2.		
DR	PROSITE; PS01286; FA58C_2; 2.		
KW	EGF-like domain; Alternative splicing; Signal; Developmental protein;		
FT	Cell adhesion; Repeat; Vascularization.		
FT	SIGNAL 1 16		
FT	POTENTIAL.		

FT CHAIN 17 480 INTEGRIN-BINDING PROTEIN DELL.
FT DOMAIN 26 59 EGF-LIKE 1.
FT DOMAIN 78 116 EGF-LIKE 2.
FT DOMAIN 123 154 EGF-LIKE 3.
FT DOMAIN 161 311 F5/8 TYPE C 1.
FT DOMAIN 322 473 F5/8 TYPE C 2.
FT SITE 96 98 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 97 98 BY SIMILARITY.
FT SITE 98 37 BY SIMILARITY.
FT SITE 31 48 BY SIMILARITY.
FT SITE 50 59 BY SIMILARITY.
FT SITE 78 89 BY SIMILARITY.
FT SITE 83 105 BY SIMILARITY.
FT SITE 107 116 BY SIMILARITY.
FT SITE 158 314 BY SIMILARITY.
FT SITE 301 305 BY SIMILARITY.
FT SITE 319 476 BY SIMILARITY.
FT SITE 66 66 A -> G (IN SHORT ISOFORM).
FT SITE 67 76 MISSING (IN SHORT ISOFORM).
FT SITE 76 76 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 480 AA; 53765 MW; F7171E23A309FD48 CRC64;

Query Match 10.28; Score 355; DB 4; Length 480;
Best Local Similarity 43.68; Pred. No. 1.2e-20;
Matches 75; Conservative 27; Mismatches 66; Indels 4; Gaps 2;

Qy 158 LSTSLFTFTKSGCYGTLCMESGVIAADPOITASSVLEWTDHTGQENSWKPKKARLKKPGP- 216
Db 307 LRMEELGCELSCSEPLGKSGHIOYQITASSIFRTLM--DMFTWEPKARLDKQGV 364

Qy 217 -PWAAFAFDEYQWQIDLNKKEKKTGTTGTTGTVVHNYVSAYRILYSDGQKWTYVRE 275
Db 365 NAWTSGHNDQSQWLQVLLVPTKVTGIITQAKDFGHVQFVGSYKLAISNDGEHMTVYQD 424

Qy 276 PGVEQDKIFQGNKDYHODVRNFPPIIARFIRNPTQWQOKIAMKMLLGC 327
Db 425 EKQRDKVQFGNDFNDTHRRKNVDPPIYARHILPWSWYGRITLRSELGCG 476

RESULT 13
O35474 PRELIMINARY; PRT; 480 AA.
AC O35474; O35475;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Integrin-binding protein DELL precursor.
GN EDL3 OR DELL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RC TISSUE=EMBRYO;
RX MEDLINE=98083109; PubMed=9420328;
RA Hidai C., Zupancic T., Penta K., Mikhail A., Kawana M.,
RA Quettermous E.E., Aoka Y., Fukagawa M., Matsui Y., Platika D.,
RA Auerbach R., Hogan B.L.M., Shodgrass R., Quettermous T.;
RT "Cloning and characterization of developmental endothelial locus-1: an
RT embryonic endothelial cell protein that binds the alphavbeta3 integrin
RT receptor.";
RL Genes Dev. 12:21-33(1998).

CC -1- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
CC INTERACTION WITH THE ALPHA-V BETA3 INTEGRIN RECEPTOR. INHIBITS
CC FORMATION OF VASCULAR-LIKE STRUCTURES. MAY BE INVOLVED IN
CC REGULATION OF VASCULAR MORPHOGENESIS OR REMODELING IN EMBRYONIC
CC DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: LONG ISOFORM (SHOWN HERE) AND
CC SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ANGIOBLASTS AND EARLY ENDOTHELIAL
CC CELLS. BY EMBRYONIC DAY 13.5, ALSO EXPRESSED IN A RESTRICTED GROUP
CC OF NON-ENDOTHELIAL CELLS INCLUDING CHONDROCYTES AND RETINAL
CC NEURONS.

CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE EMBRYO FROM DAY 7. AFTER DAY
CC 15.5, EXPRESSION DECREASES AND DISAPPEARS COMPLETELY BY THE TIME
CC OF BIRTH.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC EMBL; AF031524; AAB86585.1; -.
CC EMBL; AF031525; AAB86586.1; -.
CC HSSP; P00750; 1TPG.
CC MCD; MGI:1329025; Edil3.
CC InterPro; IPR000152; Asx_hydroxyl.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR000421; FA58_C.
CC Pfam; PF00008; EGF_3.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00001; EGF_like; 2.
CC SMART; SM00231; FA58C; 2.
CC PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
CC PROSITE; PS00022; EGF_1; UNKNOWN_3.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS01187; EGF_CA; 1.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC EGF-like domain; Alternative splicing; Signal; Developmental protein;
CC Cell adhesion; Repeat; Vascularization.
KW SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 480 INTEGRIN-BINDING PROTEIN DELL.
FT DOMAIN 26 59 EGF-LIKE 1.
FT DOMAIN 78 116 EGF-LIKE 2.
FT DOMAIN 123 154 EGF-LIKE 3.
FT DOMAIN 161 311 F5/8 TYPE C 1.
FT DOMAIN 322 473 F5/8 TYPE C 2.
FT SITE 96 98 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 97 98 BY SIMILARITY.
FT SITE 98 37 BY SIMILARITY.
FT SITE 31 48 BY SIMILARITY.
FT SITE 50 59 BY SIMILARITY.
FT SITE 78 89 BY SIMILARITY.
FT SITE 83 105 BY SIMILARITY.
FT SITE 107 116 BY SIMILARITY.
FT SITE 158 314 BY SIMILARITY.
FT SITE 301 305 BY SIMILARITY.
FT SITE 319 476 BY SIMILARITY.
FT SITE 218 221 INLQ -> VTVG (IN SHORT ISOFORM).
FT SITE 222 480 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 480 AA; 53740 MW; 4CD91EFE9261714D CRC64;

Query Match 10.18; Score 352; DB 11; Length 480;
Best Local Similarity 43.68; Pred. No. 2e-20;
Matches 75; Conservative 27; Mismatches 66; Indels 4; Gaps 2;

Qy 158 LSTSLFTFTKSGCYGTLCMESGVIAADPOITASSVLEWTDHTGQENSWKPKKARLKKPGP- 216
Db 307 LRMEELGCELSCSEPLGKSGHIOYQITASSIFRTLM--DMFTWEPKARLDKQGV 364

Qy 217 -PWAAFAFDEYQWQIDLNKKEKKTGTTGTTGTVVHNYVSAYRILYSDGQKWTYVRE 275
Db 365 NAWTSGHNDQSQWLQVLLVPTKVTGIITQAKDFGHVQFVGSYKLAISNDGEHMTVYQD 424

Qy 276 PGVEQDKIFQGNKDYHODVRNFPPIIARFIRNPTQWQOKIAMKMLLGC 327
Db 425 EKQRDKVQFGNDFNDTHRRKNVDPPIYARHILPWSWYGRITLRSELGCG 476

RESULT 14
O43737 PRELIMINARY; PRT; 2224 AA.
ID O43737
AC O43737; 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Factor V.

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Insectivora; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA BIRD C.;
RP Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC EMBL; Z95972; CAB16748.1; -.
DR HSP; P12259; ICDT.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR000421; FA58_C.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
SQ SEQUENCE 2224 AA; 251673 MW; 1013108D49690EAB CRC64;

Query Match 10.1%; Score 351; DB 4; Length 2224;
Best Local Similarity 46.4%; Pred. No. 2.6e-19;
Matches 77; Conservative 21; Mismatches 58; Indels 10; Gaps 3;

QY 166 KTSCTGLMESGVIAADPQITASSVLE--WTDHTGOENSWKPKKRLKPG--PPWAAF 221
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2062 EVNGCSTPLGMEKNGKQITASSFKKSWGDY-----WEPFRARLNAGGRVNAQAK 2115

QY 222 ATDEYQWLOLDLNKEKKTGIITGSTWVEHNYVSAYRILYDDGOKVTYREPGVEQD 281
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2116 ANNKKQLEIDLLKIKITAITITGCKSLSEMTYKSTIHYSEGVKWPYRLKSSMD 2175

QY 282 KIFGCKDYHODVRNFPPIIARFIRVNPQWQOKIAMKMLGCG 327
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2176 KIFGNTKGVKNFNPPIISIRFIRVNPQWQOKIAMKMLGCG 2221

RESULT 15
Q9WTS3 PRELIMINARY; PRT; 426 AA.
AC Q9WTS3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Milk fat globule glycoprotein MFG-E8 S (Similar to milk fat globule-
DE EGF factor 8 protein).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA BIRD C.;
RP Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC EMBL; AB025280; BAA76386.1; -.
DR EMBL; BC003904; AAH03904.1; -.
DR EMBL; BC003892; AAH03892.1; -.
DR HSP; P00740; 1EDM.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR000421; FA58_C.
DR InterPro: IPR001092; HLH_basic.
```

```
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR PRINTS; PR00010; EGFBLD.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 426 AA; 47197 MW; B182F4A20629881A CRC64;

Query Match 9.3%; Score 323.5; DB 11; Length 426;
Best Local Similarity 25.9%; Pred. No. 3.8e-18;
Matches 110; Conservative 54; Mismatches 155; Indels 105; Gaps 14;

QY 5 LLLLLLLLLLLELDAGAQKQKYGKGLGLOMHS--IESKGNIEITLFLMSGI-----HVSGRG 57
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7 LAALCGMLLCASGLFAASGDFCDSSLCNLGCTCLTGQDNDIYCLCPGFTGLVCNTERG 66

QY 58 FLASYSVIDKDLITCLDTASNFLEPEFSKYP-----ACGLLPFAEISGTIPHG- 107
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
67 PCSPNPCYNDAKCLIVTLDTQRGDIFTEYICQCPVSGIHCECTCSTQLGMEGAIADSQ 126

QY 108 -----YR-----DSSP-----LCMAGV----- 119
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
127 ISASSVYMGFMGLQRMGPGLARLYRTGIVNAWTASNYDSRKPWQVNLKRMVSGVMTQG 186

QY 120 --HAG-----VVSNTLGGQISWVI---SKGIPYESSLANN----- 150
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
187 ASRAGRAYELKTFKVAISLDGRKEFTQDESGDKFELGNLDNLSLVNMPNPTLEAQYI 246

QY 151 -VTSVVGH----LSTSLFTKTSCTGLMESGVIAADPQITASSVLEWTDHTGOENSWK 205
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
247 RLYPVSCHRGCTLRFELGCELHGCSEPLGLKNNITIPDSOMASS--SYKTWNLRFAFGW 304

QY 206 PKARLKKPGP--PWAAPATDEYQWLOLDLNKEKKTGIITGSTWVEHNYVSAYRILY 263
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
305 PHLGRLDNQGKINAWTAQNSAKEMQLQVDLTQRTQRTQRTQRTQRTQRTQRTQRTQRT 364

QY 264 SDGOKWTVYREPGVEQDKIFQGNKDYHQDVRRNFPPIIARFIRVNPQWQOKIAMKME 323
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
365 SDGQVQWTVYEEQG--SSKVFQGNLDNNSHKKNIFEKPFMARYVYVLPVSHNRLRL 422

QY 324 LLGC 327
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
423 LLGC 426
```

Search completed: January 21, 2003, 09:53:51
Job time : 30.0427 secs

Db 2050 WNGDY-----WEPLATLHNGGRVANANQAKANNKONTQVLDLTKKVTATYTGCKSL 2103
QY 251 EHNYSVARIILYSDDGKWYVREPCVQDKIFQGNKDYHODVNNFLPPIIARIRVN 310
Db 2104 SEMYKVSISIOYDOGVAMKPYRQKSSWVDKIFEGNSNTKGMNFNPPIISRIRII 2163
QY 311 PTONOQKIAMKELLGC 327
Db 2164 PXTWMSIALRLELFGC 2180

RESULT 2

US-08-480-229C-10
; Sequence 10, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Bridgid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION/DOCKET NUMBER: 28,462
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-229C-10

Query Match 10.1%; Score 352; DB 2; Length 480;
Best Local Similarity 43.6%; Pred. No. 5.8e-23;
Matches 75; Conservative 27; Mismatches 66; Indels 4; Gaps 2;

QY 158 LSTSLFTFKTSGCYGTGMEGVADPQTASSVLEWTDHTGQENSMKPKKARLKRP- 216
Db 307 LRMEILGCELSGSEPLGMSGHIQDQITASSVFRILNM--DMFTWEPKARLDKQGV 364
QY 217 -PMAAFATDEYQMLQIDLNKEKKTITGTTGSTMVEHNYSVARIILYSDDGKWYVRE 275
Db 365 NAWTSGHNDQSOMLOVLDLVPFTKVTGTTGAKDFGHVQFVGSYKLAYSNDEHMMVHOD 424
QY 276 PGVBDKIFQGNKDYHODVNNFLPPIIARIRVNPTOMOOKIAMKELLGC 327
Db 425 EKORKDVFGNFDNDTHRKNVIDPPIYARIRILPMSWIGRITLRSELLGC 476

RESULT 3
US-08-659-235C-10
; Sequence 10, Application US/08659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Bridgid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,235C
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION/DOCKET NUMBER: 28,462
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-659-235C-10

Query Match 10.1%; Score 352; DB 2; Length 480;
Best Local Similarity 43.6%; Pred. No. 5.8e-23;
Matches 75; Conservative 27; Mismatches 66; Indels 4; Gaps 2;

QY 158 LSTSLFTFKTSGCYGTGMEGVADPQTASSVLEWTDHTGQENSMKPKKARLKRP- 216
Db 307 LRMEILGCELSGSEPLGMSGHIQDQITASSVFRILNM--DMFTWEPKARLDKQGV 364
QY 217 -PMAAFATDEYQMLQIDLNKEKKTITGTTGSTMVEHNYSVARIILYSDDGKWYVRE 275
Db 365 NAWTSGHNDQSOMLOVLDLVPFTKVTGTTGAKDFGHVQFVGSYKLAYSNDEHMMVHOD 424
QY 276 PGVBDKIFQGNKDYHODVNNFLPPIIARIRVNPTOMOOKIAMKELLGC 327
Db 425 EKORKDVFGNFDNDTHRKNVIDPPIYARIRILPMSWIGRITLRSELLGC 476

RESULT 4
US-08-480-229C-14
; Sequence 14, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quertermous, Thomas
; APPLICANT: Hogan, Bridgid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; NUMBER OF SEQUENCES: 29


```
;
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: CRECC-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 943-1931
; TELEFAX: (510) 943-1189
; TELEX: N.A.
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE:
;
US-07-607-538C-4

Query Match          10.1%; Score 351; DB 1; Length 218;
Best Local Similarity 46.4%; Pred. No. 2.1e-23;
Matches 77; Conservative 21; Mismatches 58; Indels 10; Gaps 3;

QY 166 KTGCGYGLMGESGVADPOITASSVLE--WTDTGGOENSWKPKARLKKPG--PPWAAF 221
DB 56 EVNGCSTPLGMENKGIENKQITASSFKKSWMGDY-----WEPFRARLNAGGRVNAQAK 109
QY 222 ATDEYQWLQIDLNKKKKTGIITGTGTMVEHNYVSAYRILYSDGOKWTYVIREPGVEQD 281
DB 110 ANNKKQWLEIDLKIKKTAITITQGCCKSLSEMYVKSYTHYSEQGVENKPYRLKSSMVD 169
QY 282 KIFGKNKYHQDYVRNFPPIIARFIRVNPQWQOKIAMKMLGCG 327
DB 170 KIFEGNTKGVKNFNPPIISRFIRVIRPKTNQSIARLRLFLGC 215

RESULT 7
US-08-162-402B-4
; Sequence 4, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE:
;
US-07-607-538C-4

Query Match          10.1%; Score 351; DB 1; Length 218;
Best Local Similarity 46.4%; Pred. No. 2.1e-23;
Matches 77; Conservative 21; Mismatches 58; Indels 10; Gaps 3;

QY 166 KTGCGYGLMGESGVADPOITASSVLE--WTDTGGOENSWKPKARLKKPG--PPWAAF 221
DB 56 EVNGCSTPLGMENKGIENKQITASSFKKSWMGDY-----WEPFRARLNAGGRVNAQAK 109
QY 222 ATDEYQWLQIDLNKKKKTGIITGTGTMVEHNYVSAYRILYSDGOKWTYVIREPGVEQD 281
DB 110 ANNKKQWLEIDLKIKKTAITITQGCCKSLSEMYVKSYTHYSEQGVENKPYRLKSSMVD 169
QY 282 KIFGKNKYHQDYVRNFPPIIARFIRVNPQWQOKIAMKMLGCG 327
DB 170 KIFEGNTKGVKNFNPPIISRFIRVIRPKTNQSIARLRLFLGC 215

RESULT 7
US-08-162-402B-4
; Sequence 4, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE:
;
US-07-607-538C-4
```

```
;
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
;
US-08-162-402B-4

Query Match          10.1%; Score 351; DB 2; Length 218;
Best Local Similarity 46.4%; Pred. No. 2.1e-23;
Matches 77; Conservative 21; Mismatches 58; Indels 10; Gaps 3;

QY 166 KTGCGYGLMGESGVADPOITASSVLE--WTDTGGOENSWKPKARLKKPG--PPWAAF 221
DB 56 EVNGCSTPLGMENKGIENKQITASSFKKSWMGDY-----WEPFRARLNAGGRVNAQAK 109
QY 222 ATDEYQWLQIDLNKKKKTGIITGTGTMVEHNYVSAYRILYSDGOKWTYVIREPGVEQD 281
DB 110 ANNKKQWLEIDLKIKKTAITITQGCCKSLSEMYVKSYTHYSEQGVENKPYRLKSSMVD 169
QY 282 KIFGKNKYHQDYVRNFPPIIARFIRVNPQWQOKIAMKMLGCG 327
DB 170 KIFEGNTKGVKNFNPPIISRFIRVIRPKTNQSIARLRLFLGC 215

RESULT 8
US-08-162-402B-14
; Sequence 14, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 160 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
;
US-08-162-402B-14

Query Match          9.9%; Score 345; DB 2; Length 160;
Best Local Similarity 46.6%; Pred. No. 4.4e-23;
```

[illegible]

RESULT 9
 US-07-607-538C-3
 ; Sequence 3, Application US/07607538C
 ; Patent No. 5455031
 ; GENERAL INFORMATION:
 ; APPLICANT: Ceriani Dr., Roberto L.
 ; APPLICANT: Peterson Dr., Jerry A.
 ; APPLICANT: Larocca, David J.
 ; TITLE OF INVENTION: POLYPEPTIDE WITH 46
 ; TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
 ; TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
 ; TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-
 ; TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
 ; TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
 ; TITLE OF INVENTION: USE THEREOF
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: V. Amzel & Assoc.
 ; STREET: 2055 No. 5455031th Broadway
 ; CITY: Walnut Creek
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94596
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/607,538C
 ; FILING DATE: 01-NOV-1990
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Viviana Amzel
 ; REGISTRATION NUMBER: 30,930
 ; REFERENCE/DOCKET NUMBER: CRFCC-004
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (510) 943-1931
 ; TELEFAX: (510) 943-1189
 ; TELEX: N.A.

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE:
;
; US-07-607-538C-3

```

QY 216 -PWAARFATDEYQWQLDNLNKKKITGLITGSGTNVHNYVSAYRILYSDGQKWTYVR 274
 Db 105 -FNAWVAGSYGNDQWLQVDLGSSEKVTGLITQCARNFSGVQFVASYKVAYSNDSANWTEYQ 164
 QY 275 EPGVEQDKIFOGNKDYHODVNNFLPPIARFIRVNPQWQOKTAMKMELLCG 327
 Db 165 DPTGSSKIFPGNDWNHSHKKNLFTPLARYRILPVAWHNRIRALRLELCG 217

RESULT 10
 US-08-162-402B-3
 ; Sequence 3, Application US/08162402B
 ; Patent No. 5972337
 ; GENERAL INFORMATION:
 ; APPLICANT: CERIANT, ROBERTO L.
 ; APPLICANT: PETERSON, JERRY A.
 ; APPLICANT: LAROCCA, DAVID J.
 ; TITLE OF INVENTION: GLOBULIN HUMAN MILK FAT
 ; TITLE OF INVENTION: GLOBULE (HMEG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pretty, Schroeder & Poplawski
 ; STREET: 444 South Flower St., 19th Floor
 ; CITY: Los Angeles
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/162.402B
 ; FILING DATE: 03-DEC-1993
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Amzel, Viviana
 ; REGISTRATION NUMBER: 30,930
 ; REFERENCE/DOCKET NUMBER: P66 38215
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 213-622-7700
 ; TELEFAX: 213-489-4210
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 217 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; US-08-162-402R-3

40 ROUTE 60, BOWLING GREEN, OHIO 43402
 TITLE OF INVENTION: GLOBULE (HMF) ANTIGEN, FRAGMENTS & FUSION PROTEIN
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pretty, Schroeder & Poplawski
 STREET: 444 South Flower St., 19th Floor
 CITY: Los Angeles

```

; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 465 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-162-402B-8

Query Match          9.9%; Score 343; DB 2; Length 465;
Best Local Similarity 41.0%; Pred. No. 3 5e-22;
Matches 71; Conservative 28; Mismatches 68; Indels 6; Gaps 3;

QY 158 LSTSLFTFTKSGCYGTILGMSGVIADPQITASSYLE-WTQHTGQENSGWKPKKARLKPG- 215
      | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :|
Db 296 LRFELLGCELNGCANPLGLKNNNSIPDKQITASSSYKTWGLHL---FSWNPYSARLDKQGN 352

QY 216 -PPWAFATDEYQWLQIDLNKKEKITGIIITGSTMVHNYYVSAYRILYSDGQKWTYVR 274
      | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :|
Db 353 FNAWAGSYGNDQWLQVLDGSSKEVTGIITQGARFNGSVQVASYKYKWAYSNDSANWTEYQ 412

QY 275 EPGVEQDKIFGKNKDYHODYRNNFLPPIIARFIRVNPQWOKTAMKMLLGC 327
      | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :| | : :|
Db 413 DPTGSSKIFPGNDNHSKKNLFETPILARYVRLPVAWHNRLALRLELLGC 465

RESULT 15
US-08-162-402B-12
; Sequence 12, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LARocca, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMF) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930

```

Search completed: January 21, 2003, 09:56:09
Job time : 13.3951 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:56:10 : Search time 15 2974 Seconds
(without alignments)
4581.302 Million cell updates/sec

Title: US-10-060-830-3

Perfect score: 729

Sequence: 1 MPFLPLLLLLVLLLLLEDAGA.....TQEVSGAGRDGECDFKEIL 729.

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR-73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	1.6	1807	JC6319	integrin beta-4 ch
2	10	1.4	546	B32688	beta-galactosidase
3	10	1.4	629	A46500	Ly-9.2 antigen - m
4	10	1.4	677	A32611	beta-galactosidase
5	9	1.2	276	G64584	conserved hypotet
6	9	1.2	379	T11375	ubiquinol-cytochro
7	9	1.2	391	A43356	cathepsin E (EC 3.
8	9	1.2	573	JC4335	anti-mullerian hor
9	9	1.2	700	S12053	protein-tyrosine-p
10	9	1.2	728	E83228	hypothetical prote
11	9	1.2	732	IJCHCB	B-cadherin precurs
12	9	1.2	1197	T30581	neural cell adhesi
13	8	1.1	95	C84649	hypothetical prote
14	8	1.1	95	S46537	pathogen-inducible
15	8	1.1	102	F87993	protein 2C334.3 [i
16	8	1.1	104	H90197	hypothetical prote
17	8	1.1	114	A55010	neutrophil-activat
18	8	1.1	114	S45541	hypothetical prote
19	8	1.1	125	T27519	hypothetical prote
20	8	1.1	161	T32027	hypothetical prote
21	8	1.1	201	H71059	hypothetical prote
22	8	1.1	222	AC2397	ATP-binding protei
23	8	1.1	236	T14169	hypothetical prote
24	8	1.1	258	C86541	CT105 hypothetical
25	8	1.1	258	D72082	hypothetical prote
26	8	1.1	353	T07998	cytochrome c-type
27	8	1.1	353	T08001	cytochrome c-type
28	8	1.1	356	G01447	GP36b glycoprotein
29	8	1.1	358	JQ1278	histamine H2 recep

30	8	1.1	362	2	T35382	probable aspartate
31	8	1.1	396	2	A34401	cathepsin E (EC 3.
32	8	1.1	460	2	B83552	probable two-compo
33	8	1.1	477	2	T05202	pectinesterase hom
34	8	1.1	477	2	D70546	hypothetical prote
35	8	1.1	480	2	A31589	carboxypeptidase C
36	8	1.1	481	2	C95920	hypothetical membr
37	8	1.1	484	2	S60050	vacuolar processin
38	8	1.1	491	1	IJBQCP	P-cadherin - bovin
39	8	1.1	536	2	C96652	protein P23N19.7 [
40	8	1.1	576	1	ACFFA2	nicotinic acetylch
41	8	1.1	583	2	T01470	diphosphate-fructo
42	8	1.1	652	2	D85044	hypothetical prote
43	8	1.1	728	2	B71009	hypothetical prote
44	8	1.1	764	2	I48882	thyrotropin recept
45	8	1.1	767	2	S47639	acetylcholinestera

ALIGNMENTS

RESULT 1

JC6319

Integrin beta-4 chain precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 14-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jan-2000

C:Accession: JC6319

R:Feltri, M.L.; Arona, M.; Scherer, S.S.; Wrabetz, L.

Gene 186, 299-304, 1997

A:Title: Cloning and sequence of the cDNA encoding the beta4 integrin subunit in

A:Reference number: JC6319; MUID:97228432; PMID:9074510

A:Accession: JC6319

A:Molecule type: mRNA

A:Residues: 1-1807 <FEL>

A:Cross-references: GB:U60096; NID:g1401302; PIDN:AA353094.1; PID:g1401303

A:Experimental source: peripheral nerve

C:Comment: This protein belongs to a family of transmembrane receptors that media

C:Superfamily: integrin beta-4 chain; EGF homology; fibronectin type III repeat h

C:Keywords: glycoprotein; nerve; phosphoprotein; transmembrane protein

F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-1807/Product: beta4 integrin #status predicted <MAT>

F:544-574/Domain: EGF homology <EGF>

F:713-735/Domain: transmembrane #status predicted <TM>

F:327,492,580,619,697/Binding site: carbohydrate (Asn) (covalent) #status predict

F:1658/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 1.6%; Score 12; DB 2; Length 1807;

Best Local Similarity 100.0%; Pred. No. 0.0079;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FLLLLVLLLLL 15

DB 722 FLLLLVLLLLL 733

RESULT 2

B32688

beta-galactosidase-related protein - human

C:Species: Homo sapiens (man)

C:Date: 29-Jun-1990 #sequence_revision 29-Jun-1990 #text_change 21-Jul-2000

C:Accession: B32688

R:Morreau, H.; Galjart, N.J.; Gillemans, N.; Willemsen, R.; van der Horst, G.T.J.

J. Biol. Chem. 264, 20655-20663, 1989

A:Title: Alternative splicing of beta-galactosidase mRNA generates the classic. l3

A:Reference number: A32688; MUID:90062209; PMID:2511208

A:Accession: B32688

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-546 <MOR>

A:Cross-references: GB:M27508; NID:g179420; PIDN:AAA35599.1; PID:g179421

C:Keywords: alternative splicing

Query Match 1.4%; Score 10; DB 2; Length 546;

Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLVLLLL 14
| | | | |
Db 9 LLLLLVLLLL 18

RESULT 3

A46500
Ly-9.2 antigen - mouse
C:Species: Mus musculus (house mouse)
C:Date: 18-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A46500
J:Sandrin, M.S.; Gumley, T.P.; Henning, M.M.; Vaughan, H.A.; Gonsz, L.J.; Trapani, J.A.;
R:Immunol. 149, 1636-1641, 1992
A:Title: Isolation and characterization of cDNA clones for mouse Ly-9.
A:Reference number: A46500; MUID:92373005; PMID:1506686
A:Accession: A46500
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-629 <SAN>
A:Cross-references: GB:M84412; NID:g198931; PIDN:AAA39468.1; PID:g198932
A:Experimental source: C57BL/6
A:Note: sequence extracted from NCBI backbone (NCBIN:111651, NCBIIP:111654)
C:Keywords: transmembrane protein

Query Match 1.4%; Score 10; DB 2; Length 629;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLVLLLL 14
| | | | |
Db 435 LLLLLVLLLL 444

RESULT 4

A32611
beta-galactosidase (EC 3.2.1.23) precursor - human
N:Alternate names: lactase
C:Species: Homo sapiens (man)
C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 29-Sep-1999
C:Accession: A32688; B37086; A31673
J:Morreau, H.; Galjart, N.J.; Gillemans, N.; Willemsen, R.; van der Horst, G.T.J.; d'Azz
R:Biochem. Biophys. Res. Commun. 264, 20655-20663, 1989
A:Title: Alternative splicing of beta-galactosidase mRNA generates the classic lysosomal
A:Reference number: A32688; MUID:90062209; PMID:2511208
A:Accession: A32688
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-677 <MOR>
A:Cross-references: GB:M27507; NID:g179400; PIDN:AAA51819.1; PID:g179401
R:Iyanamoto, Y.; Hake, C.A.; Martin, B.M.; Kretz, K.A.; Ahern-Rindell, A.J.; Naylor, S.L.
DNA Cell Biol. 9, 119-127, 1990
A:Title: Isolation, characterization, and mapping of a human acid beta-galactosidase cDN
A:Reference number: A32611; MUID:90262647; PMID:2111707
A:Accession: A32611
A:Molecule type: mRNA
A:Residues: 1-677 <YAM>
A:Cross-references: EMBL:M34423; NID:g179422; PIDN:AAA51823.1; PID:g179423
R:Nanba, E.; Suzuki, K.
Biochem. Biophys. Res. Commun. 173, 141-148, 1990
A:Title: Molecular cloning of mouse acid beta-galactosidase cDNA: sequence, expression c
A:Reference number: A37086; MUID:91076843; PMID:2124109
A:Accession: B37086
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-677 <NAN>
R:Oshima, A.; Tsuji, A.; Nagao, Y.; Sakuraba, H.; Suzuki, Y.
Biochem. Biophys. Res. Commun. 157, 238-244, 1988
A:Title: Cloning, sequencing, and expression of cDNA for human beta-galactosidase.
A:Reference number: A31673; MUID:89061717; PMID:3143362

A:Accession: A31673
A:Molecule type: mRNA
A:Residues: 1-9, P', 11-200, A', 202-677 <OSH>
A:Cross-references: GB:M22590; NID:g179418; PIDN:AAA51822.1; PID:g179419
A:Experimental source: placenta
C:Comment: This enzyme is deficient in GM-1 gangliosidosis and Morquio B syndrome.
C:Genetics:
A:Gene: GDB:GLB1
A:Cross-references: GDB:119987; OMIM:230500
A:Map position: 3p21.33-3p21.33
C:Superfamily: beta-galactosidase bga
C:Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-677/Product: beta-galactosidase #status predicted <MAT>
F:26,247,464,498,542,545,555/Binding site: carbohydrate (Asn) (covalent) #status pr

Query Match 1.4%; Score 10; DB 2; Length 677;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLVLLLL 14
| | | | |
Db 9 LLLLLVLLLL 18

RESULT 5

G64584
conserved hypothetical protein HP0519 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: G64584
R:Romb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann,
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; Mcl
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthe
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: G64584
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-276 <TOM>
A:Cross-references: GB:AE000566; GB:AE000511; NID:g2313628; PIDN:AAD07587.1; PID:g2

Query Match 1.2%; Score 9; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LLLLLVLLLL 13
| | | | |
Db 16 LLLLLVLLLL 24

RESULT 6

T11375
ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - donkey mitochondrion
C:Species: mitochondrion Equus asinus (donkey)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
C:Accession: T11375
R:Xu, X.; Gullberg, A.; Arnason, U.
J. Mol. Evol. 43, 438-463, 1996
A:Title: The complete mitochondrial (mtDNA) of the donkey and mtDNA comparisons amon
A:Reference number: Z17265; MUID:97032591; PMID:8875857
A:Accession: T11375
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-379 <UX>
A:Cross-references: EMBL:X97337; NID:g1805746; PIDN:CAA66026.1; PID:g1805759
A:Experimental source: kidney
C:Genetics:
A:Genome: mitochondrion
C:Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocu
C:Keywords: electron transfer; heme; iron; metalloprotein; mitochondrion; oxidative

F;11-339/Domain: cytochrome b homology <CYB>
 F;11-209/Domain: cytochrome b6 homology <CB6>
 F;221-339/Domain: plastoquinol-plastocyanin reductase 17K protein homology <17k>
 F;83,182/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted
 F;97,196/Binding site: heme iron (His) (axial ligands) (high potential) #status predicted

Query Match 1.2%; Score 9; DB 2; Length 379;

Best Local Similarity 100.0%; Pred. No. 2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LLLLLLLLLL 14
 |||||
 Db 232 LLLLLLLLLL 240

RESULT 7

A43356
 cathepsin E (EC 3.4.23.34) precursor - guinea pig
 N;Alternate names: erythrocyte membrane aspartic proteinase; slow-moving proteinase
 C;Species: Cavia porcellus (guinea pig)
 C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
 C;Accession: A43356
 R;Ragayama, T.; Ichinose, M.; Tsukada, S.; Miki, K.; Kurokawa, K.; Koiwai, O.; Tanji, M.
 J. Biol. Chem. 267, 16450-16459, 1992
 A;Title: Gastric procathepsin E and progastricsin from guinea pig. Purification, molecular
 A;Reference number: A43356; MUID:92355614; PMID:1644829
 A;Accession: A43356
 A;Molecule type: mRNA
 A;Residues: 1-391 <RAG>
 A;Cross-references: GB:M88653; NID:g191294; PIDN:AAA37052.1; PID:g191295
 A;Note: sequence extracted from NCBI backbone (NCBIN:110763, NCBIPI:110769)
 C;Superfamily: pepsin
 C;Keywords: aspartic proteinase; hydrolase; membrane protein

Query Match 1.2%; Score 9; DB 2; Length 391;

Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 FLLLLLVLL 12
 |||||
 Db 4 FLLLLLVLL 12

RESULT 8

JC4335
 anti-mullerian hormone type II receptor precursor - human
 C;Species: Homo sapiens (man)
 C;Date: 06-Dec-1995 #sequence_revision 08-Feb-1996 #text_change 04-Mar-2000
 C;Accession: JC4335
 R;Visser, J.A.; McLuskey, A.; van Beers, T.; Weghuis, D.O.; van Kessel, A.G.; Grootegoed
 Biochem. Biophys. Res. Commun. 215, 1029-1036, 1995
 A;Title: Structure and chromosomal localization of the human anti-mullerian hormone type
 A;Reference number: JC4335; MUID:96028015; PMID:7488027
 A;Accession: JC4335
 A;Molecule type: mRNA
 A;Residues: 1-573 <VTS>
 A;Cross-references: GB:X91156; NID:g1107671; PIDN:CAA62593.1; PID:el98046; PID:g1107672
 C;Comment: This is a receptor for anti-mullerian hormone (see PIR:WPHUM). It plays a cri
 C;Genetics:

A;Gene: GDB:AMHR2

A;Cross-references: GDB:696210; OMIM:600956

A;Map position: 12q13-12q13

A;Introns: 16/3; 77/3; 141/3; 167/3; 207/2; 284/2; 322/3; 380/2; 429/3; 475/2

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
 C;Keywords: ATP; hormone receptor; transmembrane protein

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-573/Product: anti-mullerian hormone type II receptor #status predicted <MAT>

F;17-141/Domain: extracellular hormone binding #status predicted <ELB>

F;142-167/Domain: transmembrane #status predicted <TM>

F;201-512/Domain: protein kinase homology <KIN>

Query Match

Best Local Similarity 100.0%; Pred. No. 2.9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 LFLLLLLLVL 11
 |||||
 Db 154 LFLLLLLLVL 162

RESULT 9

SI2053
 protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type epsilon precursor - hum
 C;Species: Homo sapiens (man)
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C;Accession: SI2053
 R;Krueger, N.X.; Streuli, M.; Saito, H.
 EMBO J. 9, 3241-3252, 1990

A;Title: Structural diversity and evolution of human receptor-like protein tyrosi
 A;Reference number: SI2049; MUID:9106018; PMID:2170109

A;Accession: SI2053

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-700 <KRU>

A;Cross-references: GB:X54134; NID:g35791; PIDN:CAA38069.1; PID:g35792

C;Genetics:

A;Gene: GDB:PTPRE

A;Cross-references: GDB:131385; OMIM:600926

A;Map position: 10q26-10q26

C;Function:

A;Description: catalyzes the hydrolysis of peptidyl-phosphotyrosine to release p
 C;Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte comm
 C;Keywords: phosphoprotein; phosphoric monoester hydrolase; receptor; transmembr

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-700/Product: protein-tyrosine-phosphatase, receptor type epsilon #status pr

F;47-63/Domain: transmembrane #status predicted <TM>

F;78-698/Domain: leukocyte common antigen cytosolic domain homology <LAC>

F;159-383/Domain: protein-tyrosine-phosphatase homology <PTPI>

F;335/Active site: Cys (phosphocysteine intermediate) #status predicted

F;341/Binding site: substrate phosphate (Arg) #status predicted

F;630/Active site: Cys (phosphocysteine intermediate) #status predicted

F;636/Binding site: substrate phosphate (Arg) #status predicted

Query Match 1.2%; Score 9; DB 1; Length 700;

Best Local Similarity 100.0%; Pred. No. 3.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLVLL 13

|||||
 Db 56 LLLLLLVLL 64

RESULT 10

E83228
 hypothetical protein PA3339 [imported] - Pseudomonas aeruginosa (strain PA01)
 C;Species: Pseudomonas aeruginosa
 C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C;Accession: E83228
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, I
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, I
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunist

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: E83228

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-728 <SPO>

A;Cross-references: GB:AE004756; GB:AE004091; NID:g9949466; PIDN:AAG06727.1; GSP:

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA3339

Query Match

Best Local Similarity 100.0%; Pred. No. 3.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LLLVLLLLL 15
 |||||
 Db 4 LLLVLLLLL 12

RESULT 11
 IJCHCB
 B-cadherin precursor - chicken (fragment)
 N:Alternate names: K-CAM protein
 C:Species: Gallus gallus (chicken)
 C>Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
 C:Accession: A41634; A38715; S16160
 R:Sorkin, B.C.; Gallin, W.J.; Edelman, G.M.; Cunningham, B.A.
 Proc. Natl. Acad. Sci. U.S.A. 88, 11545-11549, 1991
 A:Title: Genes for two calcium-dependent cell adhesion molecules have similar structures
 A:Reference number: A41634; MUID:92107987; PMID:1763068
 A:Accession: A41634
 A:Molecule type: DNA
 A:Residues: 1-732 <SOR>
 A:CROSS-references: GB:H81894; NID:9212226; PIDN:AAA48929.1; PID:g212227
 R:Napolitano, E.W.; Venstrom, K.; Wheeler, E.F.; Reichardt, L.F.
 J. Cell Biol. 113, 893-905, 1991
 A:Title: Molecular cloning and characterization of B-cadherin, a novel chick cadherin.
 A:Reference number: A38715; MUID:91225083; PMID:2026653
 A:Accession: A38715
 A:Molecule type: mRNA
 A:Residues: 7-413, V, 415-732 <NAP>
 A:CROSS-references: GB:X58518; NID:g631113; PIDN:CAA41408.1; PID:g63114
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to
 C:Genetics:
 A:Gene: K-CAM
 A:Introns: 29/3; 81/3; 130/1; 188/3; 231/3; 293/3; 375/2; 423/1; 498/1; 571/1; 614/3; 66
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
 F:1-6/Domain: propeptide (fragment) #status predicted <PRO>
 F:6-554/Domain: extracellular #status predicted <EXT>
 F:7-732/Product: B-cadherin #status predicted <MAT>
 F:9-114/Domain: cadherin repeat homology <CR1>
 F:84-89/Region: cadherin binding #status predicted
 F:117-227/Domain: cadherin repeat homology <CR2>
 F:230-339/Domain: cadherin repeat homology <CR3>
 F:340-447/Domain: cadherin repeat homology <CR4>
 F:448-552/Domain: cadherin repeat homology <CR5>
 F:555-580/Domain: transmembrane #status predicted <TMW>
 F:581-732/Domain: intracellular #status predicted <INT>
 F:689-702/Region: serine-rich
 F:137,410/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.2%; Score 9; DB 1; Length 732;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LLLVLLLLL 15
 |||||
 Db 569 LLLVLLLLL 577

RESULT 12
 T30581
 neural cell adhesion molecule Ll.1 - zebra fish (fragment)
 C:Species: Brachydanio rerio (zebra fish)
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T30581
 R:Fongiorigi, E.; Bernhardt, R.R.; Schachner, M.
 J. Neurosci. Res. 42, 547-561, 1995
 A:Title: Zebrafish neurons express two Ll-related molecules during early axonogenesis.
 A:Reference number: Z20875; MUID:96155762; PMID:8568941
 A:Accession: T30581
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1197 <TON>
 A:CROSS-references: EMBL:X89204; NID:g1065713; PID:g1065714; PIDN:CAA61490.1

C:Superfamily: neural cell adhesion molecule Ll; fibronectin type III repeat homolc

Query Match 1.2%; Score 9; DB 2; Length 1197;
 Best Local Similarity 100.0%; Pred. No. 5.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LLLVLLLLL 14
 |||||
 Db 1064 LLLVLLLLL 1072

RESULT 13
 C84649
 hypothetical protein At2g25510 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: C84649
 R:Liu, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallo
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vent
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: C84649
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-95 <STO>
 A:CROSS-references: GB:AE002093; NID:g4432862; PIDN:AAD20710.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g25510
 A:Map position: 2

Query Match 1.1%; Score 8; DB 2; Length 95;
 Best Local Similarity 100.0%; Pred. No. 6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LFLLLLLLV 10
 |||||
 Db 7 LFLLLLLLV 14

RESULT 14
 S46537
 pathogen-inducible protein CXC750 precursor - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S46537
 R:Aufsatz, W.; Grimm, C.
 Plant Mol. Biol. 25, 229-239, 1994
 A:Title: A new, pathogen-inducible gene of Arabidopsis is expressed in an ecotype-sf
 A:Reference number: S46537; MUID:94289647; PMID:8018872
 A:Accession: S46537
 A:Molecule type: DNA
 A:Residues: 1-95 <AUF>
 A:CROSS-references: EMBL:X72022; NID:g457715; PIDN:CAA50905.1; PID:g457716
 C:Genetics:
 A:Gene: CXC750
 A:Introns: 36/1
 C:Function:
 A:Description: probably a member of the plant defense response system
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-95/Product: pathogen-inducible protein CXC750 #status predicted <MAT>

Query Match 1.1%; Score 8; DB 2; Length 95;
 Best Local Similarity 100.0%; Pred. No. 6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LFLLLLLLV 10
 |||||
 Db 11 LFLLLLLLV 18

RESULT 15

F87993
protein ZC334.3 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: F87993
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: F87993
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-102 <STO>
A:Cross-references: GB:chr_I; PIDN:CAB04964.1; PID:g3881432; GSPDB:GN00019; CESP:ZC334.3
C:Genetics:
A:Note: predicted using Genefinder
A:Gene: ZC334.3
A:Map position: 1

Query Match 1.1%; Score 8; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 LFLLLLLV 10
|||||||
Db 28 LFLLLLLV 35

Search completed: January 21, 2003, 10:00:13
Job time : 20.2974 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2003, 09:54:45 ; Search time 8.96744 Seconds
(without alignments)
3371.781 Million cell updates/sec

Title: US-10-060-830-3
Perfect score: 729
Sequence: 1 MPLFLLLLVLLLLDAGA.....TQEVSGAGRDGCDVFRIL 729

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	12	1.6	1807	1 ITB4_RAT	Q64632 rattus norv
2	10	1.4	546	1 BGAM_HUMAN	P16279 homo sapien
3	10	1.4	654	1 LY9_MOUSE	Q01965 mus musculu
4	10	1.4	677	1 BGAL_HUMAN	P16278 homo sapien
5	9	1.2	377	1 MLC1_HUMAN	Q15049 homo sapien
6	9	1.2	379	1 CYB_EQUAS	P92487 equus asinu
7	9	1.2	381	1 CYB_ANTNA	Q33800 antechinus
8	9	1.2	381	1 CYB_DASHA	Q34321 dasyurus ha
9	9	1.2	381	1 CYB_NEOLO	Q35157 neophascoga
10	9	1.2	381	1 CYB_PHATA	Q35673 phascogale
11	9	1.2	391	1 CATE_CAVPO	P25796 cavia porce
12	9	1.2	700	1 PTPE_HUMAN	P23469 homo sapien
13	9	1.2	732	1 CADL_CHICK	P33145 gallus gall
14	8	1.1	114	1 S205_HUMAN	P42830 homo sapien
15	8	1.1	114	1 FYUD_BACSU	P17616 bacillus su
16	8	1.1	242	1 SGY1_HUMAN	Q9UK85 homo sapien
17	8	1.1	353	1 CCSA_CHLRE	P48269 chlamydomon
18	8	1.1	356	1 VP36_CANFA	P49256 canis faml
19	8	1.1	356	1 VP36_HUMAN	I12907 homo sapien
20	8	1.1	358	1 HH2R_RAT	P25102 rattus norv
21	8	1.1	381	1 CYB_DASBY	Q34300 dasyuroides
22	8	1.1	381	1 CYB_DASCR	Q34302 dasyercus
23	8	1.1	396	1 CATE_MOUSE	P14091 homo sapien
24	8	1.1	397	1 CATE_MOUSE	P70269 mus musculu
25	8	1.1	425	1 FUTC_DROME	P83088 drosophila
26	8	1.1	480	1 PRTP_HUMAN	P10619 homo sapien
27	8	1.1	486	1 VPBB_ARATH	Q39044 arabidopsis
28	8	1.1	491	1 CAD3_BOVIN	P19535 bos taurus
29	8	1.1	576	1 ACH2_DROME	P17644 drosophila
30	8	1.1	764	1 TSHR_MOUSE	P47750 mus musculu
31	8	1.1	767	1 ACES_CHICK	P36196 gallus gall
32	8	1.1	829	1 CAD3_HUMAN	P22223 homo sapien
33	8	1.1	977	1 KFMS_MOUSE	P09581 mus musculu

34	8	1.1	978	1 KFMS_RAT	Q00495 rattus norv
35	8	1.1	1288	1 CTA3_HUMAN	Q9B276 homo sapien
36	8	1.1	1308	1 CTA4_HUMAN	Q9C0A0 homo sapien
37	8	1.1	1310	1 CTA4_MOUSE	Q9P47 mus musculu
38	8	1.1	1663	1 CO3_MOUSE	P01027 mus musculu
39	8	1.1	1663	1 CO3_RAT	P01026 rattus norv
40	7	1.0	37	1 LPPY_SERMA	P19337 serratia ma
41	7	1.0	47	1 LY52_ECOLI	P06963 escherichia
42	7	1.0	47	1 LY56_ECOLI	P13345 escherichia
43	7	1.0	47	1 LY57_ECOLI	Q03709 escherichia
44	7	1.0	47	1 LY58_ECOLI	P10099 escherichia
45	7	1.0	47	1 LY59_ECOLI	P15176 escherichia

ALIGNMENTS

RESULT 1
ID ITB4_RAT STANDARD; PRT; 1807 AA.
AC Q64632;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Integrin beta-4 precursor (GP150) (CD104 antigen).
GN ITGB4.

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97228432; PubMed=9074510;
RA Feltri M.L., Arona M., Scherer S.S., Wrabetz L.;
RT "Cloning and sequence of the cDNA encoding the beta 4 integrin
subunit in rat peripheral nerve.";
RL Gene 186:299-304(1997).
CC -!- FUNCTION: INTEGRIN ALPHA-6/BETA-4 IS A RECEPTOR FOR LAMININ. IT
PLAYS A CRITICAL STRUCTURAL ROLE IN THE HEMIDESMOSOME OF
EPITHELIAL CELLS.

CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-4
ASSOCIATES WITH ALPHA-6.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE FIBRONECTIN TYPE III-LIKE DOMAINS BIND BPAG1 AND
PLECTIN AND PROBABLY ALSO RECRUIT BP230.
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC -!- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; U60096; AAC53094.1;
DR HSSP; P05106; LJV2.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR002369; Integrin_B.
DR InterPro; IPR001169; Integrin_beta_C.
DR InterPro; IPR003659; Plectin-like.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00041; fn3; 4.
DR Pfam; PF00362; Integrin_B; 1.
DR Pfam; PF03160; Calx-beta; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR ProDom; PD001811; Integrin_B; 1.

```
DR SMART; SM00237; Calx_beta; 1.
DR SMART; SM00060; FN3; 4.
DR SMART; SM00187; INB; 1.
DR SMART; SM00423; PSI; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00243; INTEGRIN_BETA; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; UNKNOWN_2.
KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW Repeat; Signal.
FT CHAIN 1 27
FT CHAIN 28 1807
FT DOMAIN 28 712
FT DOMAIN 713 735
FT DOMAIN 736 1807
FT DOMAIN 129 366
FT DOMAIN 457 621
FT REPEAT 457 503
FT REPEAT 504 543
FT REPEAT 544 582
FT REPEAT 583 621
FT DOMAIN 1128 1216
FT DOMAIN 1221 1316
FT DOMAIN 1458 1547
FT DOMAIN 1624 1716
FT DISULFID 30 456
FT DISULFID 38 48
FT DISULFID 41 72
FT DISULFID 51 61
FT DISULFID 245 288
FT DISULFID 424 673
FT DISULFID 453 458
FT DISULFID 469 480
FT DISULFID 477 513
FT DISULFID 482 491
FT DISULFID 493 504
FT DISULFID 519 524
FT DISULFID 521 552
FT DISULFID 526 537
FT DISULFID 528 563
FT DISULFID 565 574
FT DISULFID 576 583
FT DISULFID 597 602
FT DISULFID 599 650
FT DISULFID 604 616
FT DISULFID 628 637
FT DISULFID 634 708
FT DISULFID 653 682
FT CARBOHYD 327 327
FT CARBOHYD 492 492
FT CARBOHYD 580 580
FT CARBOHYD 619 619
FT CARBOHYD 697 697
SQ SEQUENCE 1807 AA; 200588 MW; 2EF9BD4E345829A3 CRC64;
Query Match 1.6%; Score 12; DB 1; Length 1807;
Best Local Similarity 100.0%; Pred. No. 0.0048;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 FLLLLVLLLL 15
Db 722 FLLLLVLLLL 733
RESULT 2
BGAM_HUMAN
ID BGAM_HUMAN
AC P16279;
DT 01-AUG-1990 (Rel. 15, Created)
DE 01-AUG-1990 (Rel. 15, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Beta-galactosidase-related protein precursor (Beta-galactosidase-like
DE protein) (S-Gal) (Elastin-binding protein) (EBP).
```

```
GN GBL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PLACENTAL, PARTIAL SEQUENCE.
RC TISSUE=Testis;
RX MEDLINE=90062209; PubMed=2511208;
RA Morreau H., Galjart N.J., Gillemaus N., Willemssen R.,
RA van der Horst G.T.J., D'Azzo A.;
RT "Alternative splicing of beta-galactosidase mRNA generates the
RL classic lysosomal enzyme and a beta-galactosidase-related protein.";
RN J. Biol. Chem. 264:20655-20663(1989).
[2]
RP ELASTIN AND LAMININ BINDING DOMAIN.
RX PubMed=8383699;
RA Hinek A., Rabinovitch M., Keeley F., Okamura-Oho Y., Callahan J.;
RT "The 67-kD elastin/laminin-binding protein is related to an
RT enzymatically inactive, alternatively spliced form of
RT beta-galactosidase.";
RL J. Clin. Invest. 91:1198-1205(1993).
[3]
RP IDENTITY OF BETA-GALACTOSIDASE-RELATED PROTEIN WITH EBP.
RX PubMed=9497360;
RA Privitera S., Prody C.A., Callahan J.W., Hinek A.;
RT "The 67-kDa enzymatically inactive alternatively spliced variant of
RT beta-galactosidase is identical to the elastin/laminin-binding
RT protein.";
RL J. Biol. Chem. 273:6319-6326(1998).
[4]
RP ELASTIC-FIBER ASSEMBLY STUDIES.
RX PubMed=10841810;
RA Hinek A., Zhang S., Smith A.C., Callahan J.W.;
RT "Impaired elastic-fiber assembly by fibroblasts from patients with
RT either Morquio B disease or infantile GM1-gangliosidosis is linked to
RL a deficiency in the 67-kD spliced variant of beta-galactosidase.";
RN Am. J. Hum. Genet. 67:23-36(2000).
[5]
RP REVIEW.
RX PubMed=8922281;
RA Hinek A.;
RT "Biological roles of the non-integrin elastin/laminin receptor.";
RL Biol. Chem. 377:471-480(1996).
[6]
RP REVIEW.
RX PubMed=10571006;
RA Callahan J.W.;
RT "Molecular basis of GM1 gangliosidosis and Morquio disease, type B.
RT Structure-function studies of lysosomal beta-galactosidase and the
RT non-lysosomal beta-galactosidase-like protein.";
RL Biochim. Biophys. Acta 1455:85-103(1999).
CC -!- FUNCTION: This protein has no beta-galactosidase catalytic
CC activity, but plays functional roles in the formation of
CC extracellular elastic fibers (elastogenesis) and in the
CC development of connective tissue. Seems to be identical to the
CC elastin-binding protein (EBP), a major component of the non-
CC integrin cell surface receptor expressed on fibroblasts, smooth
CC muscle cells, chondroblasts, leukocytes, and certain cancer cell
CC types. In elastin producing cells, associates with tropoelastin
CC intracellularly and functions as a recycling molecular chaperone
CC which facilitates the secretions of tropoelastin and its assembly
CC into elastic fibers.
CC -!- SUBCELLULAR LOCATION: Localized to the perinuclear area of the
CC cytoplasm but not to lysosomes.
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms: 1/beta-galactosidase (AC P16278)
CC and 2 (shown here); are produced by alternative splicing.
CC -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: M27508; AAA35599.1; -
 DR PIR: B32688; B32688.
 DR Genew; HGNC:4298; GLB1.
 DR MIM; 230500; -
 DR InterPro; IPR001944; GH_35.
 DR Pfam; PR01301; Glyco_hydro.35; 1.
 KW Signal; Alternative splicing; Glycoprotein.
 FT SIGNAL 1 23
 FT CHAIN 24 546 BETA-GALACTOSIDASE-RELATED PROTEIN.
 FT DOMAIN 90 103 ELASTIN/LAMININ BINDING.
 FT CARBOHYD 26 26 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 546 AA; 60551 MW; 2BIA73EDAF9E966C CRC64;

Query Match 1.4%; Score 10; DB 1; Length 546;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLLLLL 14
 DB 9 LLLLLLLLLL 18

RESULT 3

LY9_MOUSE STANDARD: PRT; 654 AA.
 AC Q01965; Q9ES35; Q9ES36; Q9ES29;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE T-lymphocyte surface antigen Ly-9 precursor (Lymphocyte antigen 9)
 DE (Cell-surface molecule Ly-9).
 GN LY9 OR LY-9.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND POLYMORPHISM.
 RC STRAIN=129/SV, BALB/c, and C57BL/6; TISSUE=Spleen;
 RX MEDLINE=20424510; PubMed=10970093;
 RA Tovar V., de la Fuente M.A., Pizcueta P., Bosch J., Engel P.;
 RT "Gene structure of the mouse leukocyte cell surface molecule Ly9.";
 RL Immunogenetics 51:788-793(2000).
 RN [2]
 RP SEQUENCE OF 22-654 FROM N.A., AND SEQUENCE OF 48-59.
 RX MEDLINE=92373005; PubMed=1506686;
 RA Sandrin M.S., Gumley T.P., Henning M.M., Vaughan H.A., Goney L.J.,
 RA Trapani J.A., McKenzie I.F.C.;
 RT "Isolation and characterization of cDNA clones for mouse Ly-9.";
 RL J. Immunol. 149:1636-1641(1992).
 CC -|- FUNCTION: May participate in adhesion reactions between T
 CC lymphocytes and accessory cells by homophilic interaction.
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -|- TISSUE SPECIFICITY: LYMPHOCYTES.
 CC -|- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -|- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -|- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: AF244131; AAG14997.1; -
 DR EMBL: AF244130; AAG14996.1; -
 DR EMBL: AF246701; AAG13268.2; -
 DR EMBL: AF245117; AAG13268.2; JOINED.
 DR EMBL: AF245506; AAG13268.2; JOINED.
 DR EMBL: AF245118; AAG13268.2; JOINED.
 DR EMBL: AF245507; AAG13268.2; JOINED.
 DR EMBL: AF245508; AAG13268.2; JOINED.
 DR EMBL: AF245509; AAG13268.2; JOINED.
 DR EMBL: AF245510; AAG13268.2; JOINED.
 DR EMBL: AF246699; AAG13268.2; JOINED.
 DR EMBL: AF246700; AAG13268.2; JOINED.
 DR PIR: M84412; AAA39468.1; -
 DR HSP; P08921; LHNG.
 DR MGD; MGI:96885; Ly9.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003600; Ig_like.
 DR Pfam; PF00047; ig; 3
 DR SMART; SM00409; ig; 2.
 DR SMART; SM00410; IG-like; 3.
 KW Cell adhesion; Antigen; Signal; Transmembrane; Glycoprotein;
 KW Immunoglobulin domain; Repeat; Polymorphism.
 FT SIGNAL 1 47
 FT CHAIN 48 654 T-LYMPHOCYTE SURFACE ANTIGEN LY-9.
 FT DOMAIN 48 453 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 454 474 POTENTIAL.
 FT DOMAIN 475 654 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 48 158 IG-LIKE V-TYPE DOMAIN 1.
 FT DOMAIN 163 249 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 250 362 IG-LIKE V-TYPE DOMAIN 2.
 FT DOMAIN 363 453 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 172 242 POTENTIAL.
 FT DISULFID 178 222 POTENTIAL.
 FT DISULFID 376 445 POTENTIAL.
 FT DISULFID 382 426 POTENTIAL.
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 412 412 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 10 10 D -> G (IN LY9-1).
 FT VARIANT 14 14 G -> S (IN LY9-1).
 FT VARIANT 79 79 I -> T (IN LY9-1).
 FT VARIANT 91 91 F -> S (IN LY9-1).
 FT VARIANT 130 130 H -> Y (IN LY9-1).
 FT VARIANT 139 139 I -> T (IN LY9-1).
 FT VARIANT 362 362 P -> S.
 FT VARIANT 366 366 K -> N (IN LY9-1).
 FT VARIANT 377 377 E -> K (IN LY9-1).
 FT VARIANT 550 550 M -> I (IN LY9-1).
 FT VARIANT 592 592 G -> E (IN LY9-1).
 FT CONFLICT 283 283 F -> L (IN REF. 2).
 FT CONFLICT 499 499 T -> P (IN REF. 2).
 FT CONFLICT 560 560 V -> L (IN REF. 2).
 FT CONFLICT 647 654 TPTVNFNT -> SPYL (IN REF. 2).
 SQ SEQUENCE 654 AA; 73142 MW; 1CBBE99708AE8EE7 CRC64;

Query Match 1.4%; Score 10; DB 1; Length 654;
 Best Local Similarity 100.0%; Pred. No. 0.18;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLLLLLLL 14
 DB 456 LLLLLLLLLL 465

RESULT 4
 BGAL_HUMAN STANDARD; PRT; 677 AA.
 ID PL6278;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE beta-galactosidase precursor (EC 3.2.1.23) (Lactase) (Acid beta-galactosidase).
 GN GBL1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-Testis;
 RX MEDLINE-90062209; PubMed-2511208;
 RA Morreau H., Galfart N.J., Gillemans N., Willemsen R., van der Horst G.T.J., D'Azzo A.;
 RT "Alternative splicing of beta-galactosidase mRNA generates the classic lysosomal enzyme and a beta-galactosidase-related protein.";
 RL J. Biol. Chem. 264:20655-20663(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90262647; PubMed-2111707;
 RA Yamamoto Y., Hake C.A., Martin B.M., Kretz K.A., Ahern-Rindell A.J., Naylor S.L., Mudd M., O'Brien J.S.;
 RT "Isolation, characterization, and mapping of a human acid beta-galactosidase cDNA.";
 RL DNA Cell Biol. 9:119-127(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE-89061717; PubMed-3143362;
 RA Oshima A., Tsuji A., Nagao Y., Sakuraba H., Suzuki Y.;
 RT "Cloning, sequencing, and expression of cDNA for human beta-galactosidase.";
 RL Biochem. Biophys. Res. Commun. 157:238-244(1988).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Colon;
 RA Strausberg R.;
 RN Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RX ELASTIC-FIBER ASSEMBLY STUDIES.
 RA Hinek A., Zhang S., Smith A.C., Callahan J.W.;
 RT "Impaired elastic-fiber assembly by fibroblasts from patients with either Morquio B disease or infantile GM1-gangliosidosis is linked to deficiency in the 67-kD spliced variant of beta-galactosidase.";
 RL Am. J. Hum. Genet. 67:23-36(2000).
 RN [6]
 RP REVIEW.
 RX PubMed=10571006;
 RA Callahan J.W.;
 RT "Molecular basis of GM1 gangliosidosis and Morquio disease, type B. Structure-function studies of lysosomal beta-galactosidase and the non-lysosomal beta-galactosidase-like protein.";
 RL Biochim. Biophys. Acta 1455:85-103(1999).
 RN [7]
 RP VARIANTS MORQUIO B L-273; H-482 AND C-509, AND VARIANT GM1 C-494.
 RX MEDLINE-92026088; PubMed-1928092;
 RA Oshima A., Yoshida K., Shimamoto M., Fukuhara Y., Sakuraba H., Suzuki Y.;
 RT "Human beta-galactosidase gene mutations in morquio B disease.";
 RL Am. J. Hum. Genet. 49:1091-1093(1991).
 RN [8]
 RP VARIANTS GM1 CYS-49; THR-51 AND CYS-201.
 RX MEDLINE-91353572; PubMed-1909089;
 RA Nishimoto J., Namba E., Inui K., Okada S., Suzuki K.;
 RT "GM1-gangliosidosis (genetic beta-galactosidase deficiency):

RT identification of four mutations in different clinical phenotypes among Japanese patients.";
 RL Am. J. Hum. Genet. 49:566-574(1991).
 RN [9]
 RP VARIANTS GM1 THR-51; ARG-123; CYS-201; CYS-316 AND GLN-457.
 RX MEDLINE-91328151; PubMed-1907800;
 RA Yoshida K., Oshima A., Shimamoto M., Fukuhara Y., Sakuraba H., Yanagisawa N., Suzuki Y.;
 RT "Human beta-galactosidase gene mutations in GM1-gangliosidosis: a common mutation among Japanese adult/chronic cases.";
 RL Am. J. Hum. Genet. 49:435-442(1991).
 RN [10]
 RP VARIANT GM1 HIS-482.
 RX MEDLINE-93138608; PubMed-1487238;
 RA Mosna G., Fattore S., Tubiello G., Brocca S., Trubia M., Gianazza E., Gatti R., Danesino C., Minelli A., Piantanida M.;
 RT "A homozygous missense arginine to histidine substitution at position 482 of the beta-galactosidase in an Italian infantile GM1-gangliosidosis patient.";
 RL Hum. Genet. 90:247-250(1992).
 RN [11]
 RP VARIANTS GM1 CYS-208; ARG-578; HIS-590 AND GLY-632.
 RX MEDLINE-94027054; PubMed-8213816;
 RA Boustany R.-M., Qian W.-H., Suzuki K.;
 RT "Mutations in acid beta-galactosidase cause GM1-gangliosidosis in American patients.";
 RL Am. J. Hum. Genet. 53:881-888(1993).
 RN [12]
 RP VARIANT GM1 MET-82.
 RX MEDLINE-94256487; PubMed-8198123;
 RA Chakraborty S., Rafi M.A., Wenger D.A.;
 RT "Mutations in the lysosomal beta-galactosidase gene that cause the adult form of GM1 gangliosidosis.";
 RL Am. J. Hum. Genet. 54:1004-1013(1994).
 RN [13]
 RP VARIANTS GM1 SER-148; TYR-214; ALA-216 AND GLY-532.
 RA Hilson W.L., Okamura-Oho Y., Zhang S., Clarke J.T.R., Mahuran D., Callahan J.W.;
 RT "Novel missense mutations in beta-galactosidase that result in GM1-gangliosidosis.";
 RL Am. J. Hum. Genet. 55:A223-A223(1994).
 RN [14]
 RP VARIANTS MORQUIO B HIS-83 AND CYS-482.
 RX MEDLINE-96049832; PubMed-7586649;
 RA Ishii N., Oohira T., Oshima A., Sakuraba H., Endo F., Matsuda I., Sukegawa K., Orii T., Suzuki Y.;
 RT "Clinical and molecular analysis of a Japanese boy with Morquio B disease.";
 RL Clin. Genet. 48:103-108(1995).
 RN [15]
 RP VARIANT GM1 SER-263.
 RA Suzuki Y., Sakuraba H., Oshima A.;
 RT "Beta-galactosidase deficiency (beta-galactosidosis): GM1 gangliosidosis and Morquio B disease.";
 RL (In) Scriver C.R., Beaudet A.L., Sly W.S., Valle D. (eds.); The metabolic and molecular bases of inherited disease, pp.2787-2823, McGraw-Hill Publishing Co., New York (1995).
 RN [16]
 RP VARIANTS GM1 HIS-59; ASN-591 AND CYS-591.
 RA Morrone A., Bardeci T., Donati M.A., Giorgi M., Di Rocco R., Gatti R., Taddeucci G., Ricci R., D'Azzo A., Zammarchi E.;
 RT "Identification of new mutations in six Italian patients affected by a variant form of infantile GM1-gangliosidosis with severe cardiomyopathy.";
 RL Am. J. Hum. Genet. 61:A258-A258(1997).
 RN [17]
 RP VARIANTS GM1 HIS-201; SER-266 AND CYS-509.
 RX PubMed-9203065;
 RA Kaye E.M., Shalish C., Livermore J., Taylor H.A., Stevenson R.E., Breakfield X.O.;
 RT "Beta-Galactosidase gene mutations in patients with slowly progressive GM1 gangliosidosis.";
 RL J. Child Neurol. 12:242-247(1997).

RN [18]
 RP VARIANTS MORQUIO B GLU-438; LYS-484 AND ALA-500.
 RA Skomrowski M.-A., Bigshaw R., Zhang S., Whelan D., Clarke J.T.R.,
 RA Callahan J.W.;
 RT "Novel mutations (Asn484Lys, Thr500Ala, Gly438Glu) in Morquio B
 RT disease.";
 RL J. Inherit. Metab. Dis. 22 Suppl. 1:131-131(1999).
 RN [19]
 RP VARIANTS GMI H-59; S-121; C-208; M-240 AND N-491, AND VARIANTS P-10;
 RP C-521 AND G-532.
 RX MEDLINE=99268417; PubMed=10338095;
 RA Silva C.M.D., Severini M.H., Sopesa A., Coelho J.C., Zaha A.,
 RA D'Azio A., Giugliani R.;
 RT "Six novel beta-galactosidase gene mutations in Brazilian patients
 RT with GMI-gangliosidosis.";
 RL Hum. Mutat. 13:401-409(1999).
 RN [20]
 RP VARIANTS GMI SER-148 AND ASN-332, AND VARIANT GLY-532.
 RX PubMed=10839995;
 RA Zhang S., Bagshaw R., Hilson W., Oho Y., Hinek A., Clarke J.T.R.,
 RA Hinek A., Callahan J.W.;
 RT "Characterization of beta-galactosidase mutations Asp332-->Asn and
 RT Arg148-->Ser, and a polymorphism, Ser532-->Gly, in a case of GMI
 RT gangliosidosis.";
 RL Biochem. J. 348 3:621-632(2000).
 RN [21]
 RP VARIANTS GMI HIS-59; HIS-201; HIS-482; ASP-579; ASN-591 AND CYS-591.
 RX PubMed=10737981;
 RA Morrone A., Bardelli T., Donati M.A., Giorgi M., Di Rocco M.,
 RA Gatti R., Parini R., Ricci R., Taddeucci G., D'Azio A., Zammarchi E.;
 RT "Beta-galactosidase gene mutations affecting the lysosomal enzyme and
 RT the elastin-binding protein in GMI-gangliosidosis patients with
 RT cardiac involvement.";
 RL Hum. Mutat. 15:354-366(2000).
 RN [22]
 RP VARIANT MORQUIO B L-273, AND VARIANTS GMI M-82; H-201; D-270; Y-281;
 RP P-408 AND A-500.
 RX PubMed=11511921;
 RA Paschke E., Milos I., Kreimer-Erlacher H., Hoefler G., Beck M.,
 RA Hoeltzenbein M., Kleijer W., Levade T., Michelakakis H., Radeva B.;
 RT "Mutation analyses in 17 patients with deficiency in acid
 RT beta-galactosidase: three novel point mutations and high correlation
 RT of mutation W273L with Morquio disease type B.";
 RL Hum. Genet. 103:159-166(2001).
 CC -1- FUNCTION: Cleaves beta-linked terminal galactosyl residues from
 CC gangliosides, glycoproteins, and glycosaminoglycans.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
 CC galactose residues in beta-D-galactosides.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2/Beta-
 CC galactosidase-related protein (AC P16279); are produced by
 CC alternative splicing.
 CC -1- DISEASE: Defects in GLB1 are the cause of GMI-gangliosidosis, an
 CC autosomal recessive disorder with three major clinical phenotypes
 CC that are distinguished according to the age of onset and severity
 CC of symptoms: infantile (type I), juvenile (type II) and adult
 CC (type III). The infantile form is rapidly progressive and results
 CC Query Match 1.4%; Score 10; DB 1; Length 677;
 CC Best Local Similarity 100.0%; Pred. No. 0.19;
 CC Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 LLLLLLLLLL 14
 Db 9 LLLLLLLLLL 18
 RESULT 5
 ID MCL1_HUMAN STANDARD; PRT; 377 AA.
 AC Q15049; Q9UGY8; Q9GRP5;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Membrane protein MCL1.
 GN MCL1 OR WKL1 OR KIAA0027.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=96051387; PubMed=7584026;
 RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
 RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
 RT "Prediction of the coding sequences of unidentified human genes. I.
 RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
 RT analysis of randomly sampled cDNA clones from human immature myeloid
 RT cell line KG-1.";
 RL DNA Res. 1:27-35(1994).
 RN [2]
 RP REVISIONS.
 RA Ohara O., Nagase T., Kikuno R., Nomura N.;
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND VARIANT PC MET-309.
 RC TISSUE=Hippocampus;
 RX MEDLINE=21225990; PubMed=11326298;
 RA Meyer J., Huberth A., Ortega G., Syagailo Y.V., Jatzke S.,
 RA Moessner R., Strom T.M., Ulzheimer-Teuber I., Stoeber G., Schmitt A.,
 RA Lesch K.P.;
 RT "A missense mutation in a novel gene encoding a putative cation
 RT channel is associated with catatonic schizophrenia in a large
 RT pedigree.";
 RL Mol. Psych. 6:302-306(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Baguley M., Smith L.J., Alnsough R., Almeida J.P., Babbage A.K.,
 RA Bagnall C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chissole S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton G., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
 RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,

Qy 1195 GAGCTGCTCGGATGTCAGTTT 1215
|| || | || || | ||
Db 6739 GAACTCTTGGCTGTGATATT 6759

Search completed: January 21, 2003, 09:30:13
Job time : 114.569 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	2205	96.7	5657	9	US-09-974-298-96	Sequence 96, Appl
2	544.6	23.9	640	10	US-09-879-536-480	Sequence 480, Appl
3	488.8	21.4	580	10	US-09-864-761-12400	Sequence 12400, Appl
4	467	20.5	467	10	US-09-864-761-30344	Sequence 30344, A
5	293.2	12.9	2145	9	US-10-003-132-12	Sequence 12, Appl
6	232.4	10.2	568	10	US-09-864-761-12562	Sequence 12562, A
7	228	10.0	228	10	US-09-864-761-29163	Sequence 29163, A
8	224.2	9.8	3151	9	US-10-003-132-1	Sequence 1, Appl
9	182	8.0	1509	9	US-10-003-132-13	Sequence 13, Appl
10	149.8	6.6	2836	9	US-10-003-132-3	Sequence 3, Appl
11	148.2	6.5	1871	10	US-09-823-038A-43	Sequence 43, Appl
12	115.4	5.1	1374	9	US-10-003-132-14	Sequence 14, Appl
13	92.6	4.1	2868	9	US-10-003-132-5	Sequence 5, Appl
14	89.8	3.9	4599	9	US-09-974-298-167	Sequence 167, Appl
15	86.6	3.8	6909	10	US-09-880-107-2275	Sequence 2275, Ap
16	86.6	3.8	6925	9	US-10-115-563-13	Sequence 13, Appl
17	86.6	3.8	6925	9	US-10-115-563-26	Sequence 26, Appl
18	76.8	3.4	1270	10	US-09-880-107-3020	Sequence 3020, Ap
19	64.6	2.8	970	10	US-09-215-450-14	Sequence 14, Appl

QY 256 ATGGGAGAGAGATTGCGCATCAAAATTTGGTGACTTTTGACATTTGAAGATTTCTGATCTTGT 315
|||||
Db 181 ATGGGAGAGAGATTGCGCATCAAAATTTGGTGACTTTTGACATTTGAAGATTTCTGATCTTGT 240
QY 316 CACTTTAATTTACTTGAGAAATTTATATGGAATTTGGAGTTCAGACAGCAATGAAATAGGCAAA 375
|||||
Db 241 CACTTTAATTTACTTGAGAAATTTATATGGAATTTGGAGTTCAGACAGCAATGAAATAGGCAAA 300
QY 376 TACTGTGGTCTGGGTTGCAATGAACCAATTCAAATTTGAATCAAAAGGCAATCAATCAACA 435
|||||
Db 301 TACTGTGGTCTGGGTTGCAATGAACCAATTCAAATTTGAATCAAAAGGCAATCAATCAACA 360
QY 436 TTGCTGTTTCATGAGTGGAAATTCATGTTTCTGGACGCGGATTTTGGCCCTCATACTCTGTT 495
|||||
Db 361 TTGCTGTTTCATGAGTGGAAATTCATGTTTCTGGACGCGGATTTTGGCCCTCATACTCTGTT 420
QY 496 ATAGATAAACAAGATCTAATTTACTTGTGGACACTGCATCCAAATTTTGGAACTGAG 555
|||||
Db 421 ATAGATAAACAAGATCTAATTTACTTGTGGACACTGCATCCAAATTTTGGAACTGAG 480
QY 556 TTCAGTAAGTACTGCCAGCTGCTGCTGCTTCTGCTTCTGCTGAGATATCTGGAACAAT 615
|||||
Db 481 TTCAGTAAGTACTGCCAGCTGCTGCTGCTTCTGCTTCTGCTGAGATATCTGGAACAAT 540
QY 616 CCTCATGGATATAGAGATTCCTGCGCAATTTGTCATGGCTGGTGTGCATGACGAGTGTG 675
|||||
Db 541 CCTCATGGATATAGAGATTCCTGCGCAATTTGTCATGGCTGGTGTGCATGACGAGTGTG 600
QY 676 TCAACACAGTTGGGCGGCCAATCAGTGTGTAATTTAGTAAAGGTATTCCTTATTATGAA 735
|||||
Db 601 TCAACACAGTTGGGCGGCCAATCAGTGTGTAATTTAGTAAAGGTATTCCTTATTATGAA 660
QY 736 AGTTCTTGGCTAACACGTCACATCTGTGGGGACACTTATCTACAGTCTTTTACAA 795
|||||
Db 661 AGTTCTTGGCTAACACGTCACATCTGTGGGGACACTTATCTACAGTCTTTTACAA 720
QY 796 TTTTAAACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATCGCGATCCT 855
|||||
Db 721 TTTTAAACAAGTGGATGTTATGGAACACTGGGATGGAGTCTGGTGTGATCGCGATCCT 780
QY 856 CAATTAACAGCATCATCTGTGCTGGAGTGGAGTGCACACACAGGCAAGACAGTGG 915
|||||
Db 781 CAATTAACAGCATCATCTGTGCTGGAGTGGAGTGCACACACAGGCAAGACAGTGG 840
QY 916 AAACCCAAAAGCCAGGCTGAAAACCTGGACCCCTTGGGCTGCTTTTGGCCACTGAT 975
|||||
Db 841 AAACCCAAAAGCCAGGCTGAAAACCTGGACCCCTTGGGCTGCTTTTGGCCACTGAT 900
QY 976 GAATACCAAGTGGTTACAAATAGATTTGAATAGGAAGAAATAACAGGCAATTAAC 1035
|||||
Db 901 GAATACCAAGTGGTTACAAATAGATTTGAATAGGAAGAAATAACAGGCAATTAAC 960
QY 1036 ACTGGATCCACCATGGTGGAGCAATTTACTATGTGCTGCTTACAGAAATCTGTACAGT 1095
|||||
Db 961 ACTGGATCCACCATGGTGGAGCAATTTACTATGTGCTGCTTACAGAAATCTGTACAGT 1020
QY 1096 GATGATGGGAGAAATGGAGTGTGTACAGAGACCTGGTGTGGAGCAAGATAAGATATTT 1155
|||||
Db 1021 GATGATGGGAGAAATGGAGTGTGTACAGAGACCTGGTGTGGAGCAAGATAAGATATTT 1080
QY 1156 CAAGGAACAAGATTTATCACCAGGATGCGGTAAATCTTTTGGCCCAATTTATGCA 1215
|||||
Db 1081 CAAGGAACAAGATTTATCACCAGGATGCGGTAAATCTTTTGGCCCAATTTATGCA 1140
QY 1216 CGTTTATTAGTGAATCTTACCCTTACCAATGGCAGCAAAATTTGCCATGAAATGGAGCTG 1275
|||||
Db 1141 CGTTTATTAGTGAATCTTACCCTTACCAATGGCAGCAAAATTTGCCATGAAATGGAGCTG 1200
QY 1276 CTCGGATGTCAGTTTATTCCTAAAGTTCGCTCCTCCAAAATTTACTCAACCTCCACCTCCT 1335
|||||
Db 1201 CTCGGATGTCAGTTTATTCCTAAAGTTCGCTCCTCCAAAATTTACTCAACCTCCACCTCCT 1260
QY 1336 CGGAACAGCAATGACCTCAAAAACACTACAGGCCCTTCCAAAATTAGCCAAAGGCTGCTGCC 1395

Db 1261 CGGAACAGCAATGACCTCAAAAACACTACAGGCCCTTCCAAAATTAGCCAAAGGCTGCTGCC 1320
QY 1396 CCAAAATTTACCAACCACTACAACTCCAGTAGCAATGAATTTCTCTGCACAGAGAA 1455
|||||
Db 1321 CCAAAATTTACCAACCACTACAACTCCAGTAGCAATGAATTTCTCTGCACAGAGAA 1380
QY 1456 CAAACAACCTGCCAGTCTCTGATATCAGAAATACCTACCGTAACTCCAAATGTAACCAAGAT 1515
|||||
Db 1381 CAAACAACCTGCCAGTCTCTGATATCAGAAATACCTACCGTAACTCCAAATGTAACCAAGAT 1440
QY 1516 GTAGCGCTGGCTGCAAGTCTTGTCTGCTGCTGCTCATGGTCTCTACTACTCTCAATCTC 1575
|||||
Db 1441 GTAGCGCTGGCTGCAAGTCTTGTCTGCTGCTGCTCATGGTCTCTACTACTCTCAATCTC 1500
QY 1576 ATATTAGTGTGCTGTGGCAGTGGGACACAGAAACAGAAAGAAACTGAAGCAGCTATGAC 1635
|||||
Db 1501 ATATTAGTGTGCTGTGGCAGTGGGACACAGAAACAGAAAGAAACTGAAGCAGCTATGAC 1560
QY 1636 TTACCTTACTTGGGACCGGCGAGGTTGCTGGAAGGAAATGAAGCAGTCTTCTTCTGCAAAA 1695
|||||
Db 1561 TTACCTTACTTGGGACCGGCGAGGTTGCTGGAAGGAAATGAAGCAGTCTTCTTCTGCAAAA 1620
QY 1696 GCAGTGGACCATGAGAGAAACCCAGTTCGCTATAGCAGCAGCAGGAATTAATCACTGAGT 1755
|||||
Db 1621 GCAGTGGACCATGAGAGAAACCCAGTTCGCTATAGCAGCAGCAGGAATTAATCACTGAGT 1680
QY 1756 CCAAGAGAGTCAACACAGTGTGCGAGGCTGCTGCGAGAGTATGCTCAGCAGCTGGTA 1815
|||||
Db 1681 CCAAGAGAGTCAACACAGTGTGCGAGGCTGCTGCGAGAGTATGCTCAGCAGCTGGTA 1740
QY 1816 GGAGGAATTTGTTGGTACACTTCAACAAAGATCTACCTTTAAACCAGAAAGAAAGAA 1875
|||||
Db 1741 GGAGGAATTTGTTGGTACACTTCAACAAAGATCTACCTTTAAACCAGAAAGAAAGAA 1800
QY 1876 GCAGCTATGACAGACCTAGATCTTACAACTACCAAGGCGAGGAATTTATCATGCTAT 1935
|||||
Db 1801 GCAGCTATGACAGACCTAGATCTTACAACTACCAAGGCGAGGAATTTATCATGCTAT 1860
QY 1936 GCTGAACCACTCCCAATTTACGGGCTGAGTATGCAACCCCAATCATGAGGACATGCA 1995
|||||
Db 1861 GCTGAACCACTCCCAATTTACGGGCTGAGTATGCAACCCCAATCATGAGGACATGCA 1920
QY 1996 GGGCACCACCAACTTACGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAAC 2055
|||||
Db 1921 GGGCACCACCAACTTACGTTGGTTCAGCCCTCCACATCCACTTTCAAGGCTACGGGAAC 1980
QY 2056 CAACCTCCCCACTAGTGGGAACCTTACATACACTTCTCTCCAGGACTGACAGCTGCTCC 2115
|||||
Db 1981 CAACCTCCCCACTAGTGGGAACCTTACATACACTTCTCTCCAGGACTGACAGCTGCTCC 2040
QY 2116 TCAGCCAGGCGCCAGTATGATACCCGAAAGCTGGGAAGCCAGGCTTACCTGCCCCAGAC 2175
|||||
Db 2041 TCAGCCAGGCGCCAGTATGATACCCGAAAGCTGGGAAGCCAGGCTTACCTGCCCCAGAC 2100
QY 2176 GAATTTGGTACAGTGGCCAGCAGACACAAAGAGTATCAGAGCAGGAAGGATGGG 2235
|||||
Db 2101 GAATTTGGTACAGTGGCCAGCAGACACAAAGAGTATCAGAGCAGGAAGGATGGG 2160
QY 2236 GAATGTGATGTTTTTAAAGAAATCTTTTGAAGATGATGCTGCTTT 2280
Db 2161 GAATGTGATGTTTTTAAAGAAATCTTTTGAAGATGATGCTGCTTT 2205

RESULT 2

US-09-879-536-480/C
; Sequence 480, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.

Db

US-09-864-761-12400 Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR MICROARRAY ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 12400
LENGTH: 580
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO ACO13497.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
US-09-864-761-12400

Query Match
Best Local Similarity 99.6%; DB 10; Length 580;
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy

TCTGCAGGATGTCTCAGCCACTGGTAGGAGAATTCTTGTTCATACTTTCAAAATCT 1848
||| ||||||||||||||||||||||||||||||||||||||||||||||||||
||| ||||||||||||||||||||||||||||||||||||||||||||||||||

Db

US-09-864-761-12400 Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR MICROARRAY ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 12400
LENGTH: 580
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO ACO13497.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
US-09-864-761-12400

Query Match
Best Local Similarity 99.6%; DB 10; Length 580;
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy

TCTGCAGGATGTCTCAGCCACTGGTAGGAGAATTCTTGTTCATACTTTCAAAATCT 1848
||| ||||||||||||||||||||||||||||||||||||||||||||||||||
||| ||||||||||||||||||||||||||||||||||||||||||||||||||

Db 64 TCCACAGATGCTCAGCCACTGGTAGGAGGAATTTGGTGACACTTTCATCAAGATCT 123
QY 1849 ACCTTTAAACCAGAGAAAGAAAGCAGGCTATGACAGCTAGATCTTACAACTCA 1908
Db 124 ACCTTTAAACCAGAGAAAGAAAGCAGGCTATGACAGCTAGATCTTACAACTCA 183
QY 1909 CCAGGCGAGGAGCTTTATCATGCTCTGCTGAACCACTCCCAATACGGGGCTGAGTAT 1968
Db 184 CCAGGCGAGGAGCTTTATCATGCTCTGCTGAACCACTCCCAATACGGGGCTGAGTAT 243
QY 1969 GCAACCCCAATCATATGACATGTCAGGCGACCCCACTTCACTTCACTGAGTAT 2028
Db 244 GCAACCCCAATCATATGACATGTCAGGCGACCCCACTTCACTTCACTGAGTAT 303
QY 2029 ACATCCACTTCAAGCTAGCGGGAACCACTCCCACTTCACTTCACTTCACTTCACT 2088
Db 304 ACATCCACTTCAAGCTAGCGGGAACCACTCCCACTTCACTTCACTTCACTTCACT 363
QY 2089 CTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGGCGCCAGTATGATATCACTCCCAAGCT 2148
Db 364 CTCTCTCCAGGACTGACAGCTGCTCTCAGCCAGGCGCCAGTATGATATCACTCCCAAGCT 423
QY 2149 GGAAGCCAGGCTTACTGCTCCAGCAGCAATGCTGACAGGTCACAGGACACACAA 2208
Db 424 GGAAGCCAGGCTTACTGCTCCAGCAGCAATGCTGACAGGTCACAGGACACACAA 483
QY 2209 GAAGTATCAGGACGAGGAGGATGGGATGATGATGATGATGATGATGATGATGATGAT 2268
Db 484 GAAGTATCAGGACGAGGAGGATGGGATGATGATGATGATGATGATGATGATGATGAT 543
QY 2269 TGATGCTGCTTT 2280
Db 544 TGATGCTGCTTT 555

RESULT 4

US-09-864-761-30344
; Sequence 30344, Application US/09864761
; Patent No. US2002048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30344
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013497.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
; OTHER INFORMATION: EST_HUMAN HIT: AW583777.1, EVALUATE 3.00e-51
; OTHER INFORMATION: NT HIT: Z70177.1, EVALUATE 2.90e-01
; OTHER INFORMATION: SWISSPROT HIT: Q9Y011, EVALUATE 7.40e-01
US-09-864-761-30344
Query Match 20.5%; Score 467; DB 10; Length 467;
Best Local Similarity 100.0%; Pred. No. 2.6e-124;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1796 ACTATGCTCAGCCACTGGTAGGAGGAATTTGGTGACACTTTCATCAAGATCTACCTTTA 1855
Db 1 AGTATGCTCAGCCACTGGTAGGAGGAATTTGGTGACACTTTCATCAAGATCTACCTTTA 60
QY 1856 AACGAGAAGAGAAAGAAAGCAGGCTATGACAGCTATGACAGCTTCACTTCACTTCACTTCACTTCACT 1915
Db 61 AACGAGAAGAGAAAGAAAGCAGGCTATGACAGCTTCACTTCACTTCACTTCACTTCACTTCACT 120
QY 1916 AGGAAGTTTATCATGCTCTGCTGAACCACTCCCAATACGGGGCTGAGTATGCAACCC 1975
Db 121 AGGAAGTTTATCATGCTCTGCTGAACCACTCCCAATACGGGGCTGAGTATGCAACCC 180
QY 1976 CAATCATCATGGACATGCTCAGGCGACCCCACTTCACTTCACTTCACTTCACTTCACTTCACT 2035
Db 181 CAATCATCATGGACATGCTCAGGCGACCCCACTTCACTTCACTTCACTTCACTTCACTTCACT 240
QY 2036 CTTTCAAGGCTACGCGGAACCACTCCCACTAGTGGGAACCTTCACTTCACTTCACTTCACTTCT 2095
Db 241 CTTTCAAGGCTACGCGGAACCACTCCCACTAGTGGGAACCTTCACTTCACTTCACTTCACTTCT 300
QY 2096 CCAGGACTGACAGCTGCTCTCAGCCAGGCGCCAGTATGATATATATATATATATATATATATATAT 2155
Db 301 CCAGGACTGACAGCTGCTCTCAGCCAGGCGCCAGTATGATATATATATATATATATATATATATAT 360
QY 2156 CAGGTCTACCTGCCCCCAGACGAATTTGGTGACAGGTCCTCAGAGGACACACAGAGTAT 2215
Db 361 CAGGTCTACCTGCCCCCAGACGAATTTGGTGACAGGTCCTCAGAGGTCCTCAGAGGACACACAGAGTAT 420
QY 2216 CAGGAGCAGGAAGGATGGGGAATGTGATGTTTTTAAAGAAATCTT 2262
Db 421 CAGGAGCAGGAAGGATGGGGAATGTGATGTTTTTAAAGAAATCTT 467
RESULT 5
US-10-003-132-12
; Sequence 12, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren

APPLICANT: Shoemaker, Kimberly E.
TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUBS
FILE REFERENCE: 00-62
CURRENT APPLICATION NUMBER: US/10/003.132
BEST FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/249,004
PRIOR FILING DATE: 2000-11-15
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 2145
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: degenerate nucleotide sequence
NAME/KEY: misc_feature
LOCATION: (1)..(2145)
OTHER INFORMATION: n = A,T,C or G

US-10-003-132-12

	Query Match	12.98;	Score 293.2;	DB 9;	Length 2145;
	Best Local Similarity	35.18;	Pred. No. 7.3e-74;	Mismatches 229;	Indels 42; Gaps 4;
Qy	85 TTCTTCCGTCCTACCTGGTGCTGTCTCGTCTGCAGGAGCCGTGGAGGCCAGCAAGTT 144				
Dd	55 YTNNTNGNYTYNTHNCNGTWNSCNCNCTYNNMGNITNCARGCNGARGARTNGGN 114				
Qy	145 GATGATGTCGACACACTGTACTAGCCCTCAGAGTGGAACCTTACATCCATAACTAC 204				
Dd	115 GAYGNTGYGNCAYTYNGTNACNTAYCARGAYWSNGNACNATGACNWSNAARAAYTAY 174				
Qy	205 CCACAGACCTATCCCAACAGCACGCTTGTGTGAATGGGAGATCCGTGTAAGAATGGGAGAG 264				
Dd	175 CCGNACNATAYCNAAVCATACNTNYGARARACNATHACNGTNCNRAAGGNAR 234				
Qy	265 AGAGTTCGATCAAATTGGTGACCTTGTGACATFTGAAGATCTCATCTTGTGCTTCAAT 324				
Dd	235 MGNVTHNYTMGMNTHNGGNGNYTNGAYTAGHAR--WSNCARACNTGYGONWNGAY 291				
Qy	325 TACTTGAGAAATTATATGGAAATTCGGAGTACCAGAACCTGAATAGGCAATACTCTGTT 384				
Dd	292 TAYTNTYTNTTY-----ACNWSNWSNNGAYCARATAYGNGCNTAITYG 336				
Qy	385 CTGGGTTTTAAAATGAACCATTCATTAATCAATCAATGAATCAATCTGCTTCTTC 444				
Dd	337 GGNSNATGACNGTNCNNAARGARYTNTYNAAVCNWSNGRGTNACNGTNGMTTV 396				
Qy	445 ATGAGTGAATCAATGTTTCTGGAGCGGATTTTGGCTCATACTCTCTTTATAGATAAA 504				
Dd	397 GARWSNGWCNCAATATHSNGNGNMGNNGTYYTNTYNACNTAYTCNWSNWSNNGAYCAY 456				
Qy	505 CAAGATCTAAATCTGTTTGGACACTGCATCACTATTTTGGAACTCAGTTCAGTAAG 564				
Dd	457 CCNGAYTNTNTHACNTGYTNGARMNGCNWSCNCAATAYTNTNAAACNGARTAYWSNAR 516				
Qy	565 TACTGCCAGCTGGTGTCTGCTCTTCTGTGAGATATCTGGACAATCTCTCATGGA 624				
Dd	517 TTYTGCCGCGNGTGTGNGAYGTNGCNGNGNGAYATHSNGNGNAYATGTTNGAYGNN 576				
Qy	625 TATAGAGATTCCTGCCATTTGTCATGGCTGTGTGTCATGCGAGGAGTGTGTCACAAACAG 684				
Dd	577 TAYMGNGAYACWSNYTNTNTYAARGCNGCNATHCAIYCNGGNGNATHHGCGNGAYGAR 636				
Qy	685 TTGGCGGCCCAATCAGCTGTTGGTGGACACTTATCTACAAAGTCTTTTACATTTAAGACA 744				
Dd	637 YTNGSGNGNCARATHWSNGTNTYNCARMGNAARGGNATHWSNMGNTAYGARGGNATHYT 696				
Qy	745 GCTAACACGCTCACATCTGCTGTGGGACACTTATCTACAAAGTCTTTTACATTTAAGACA 804				
Dd	697 GCNAAYGNGNTNYTWNWSNNGAYGNGWSNYTWNWSNGAYAARMGMTTYTNTYACNWSN 756				
Qy	805 AGTGATGTTATGGAACACTGGGGATGGAGTCTGTGTGTCATGCGGATCTCAATAAACA 864				

Db 757 AYCGGNTGWSNMGNWSNTNWSTTYGARCNGAYGG-----NCARATHTGN 804

Qy 865 GCATCATCTGTCTGCTGGAGTGGACTGACACAGGAGGAGGAGCAAGTGTGGAACCCAAA 924

Dd 805 GCNWSNWSNWTGGCCARWSNGTNAAYGARWSNGNGAYCARGTNCAYTGGWSNCCNGN 864

Qy 925 AAAGCCAGGCTGAAAAAACCTGCACCGCTTGGCGCTTCTTTCCTCACTGATGAATACC--- 982

Dd 865 CARGCMNGNTNCARGAYCARGCNGCNCNWSNTGGCNGWSNGNGAYWSNWSNAAAYAAYCAY 924

Qy 983 -----AGTGGTTACAAATAGATTTCATAAAGGAAAAAGAAATAAACAGCATTTATA 1032

Dd 925 AACRCNMWGNCARTGGTYNGARATHGAYTYNGNGNGARARAARAARATHACNGNATHMGN 984

Qy 1033 ACCACTGGATCCACCATGGTGGAGCACAAATTTACTATGTCTCTGCTACAGAAATCCTGTAC 1092

Dd 985 ACNACNGNWSNACNCARWSNAAAYTVAAYTTYTAAGTNAARSNTTYGTNATCAAYTTY 1044

Qy 1093 AGTGATGATGGGAGAAATGGAGTGTACAGAGAGCCGTGTTGGGAGCAAGATAAGATA 1152

Dd 1045 AAAAAYAAAYASNARTGGAARACNTAYAAAGGNATHGCTNAAYAAYGARGAPAARGTN 1104

Qy 1153 TTTCAAGGAACAAAGATTTATCACCAGGATGCTGCTAAATACCTTTTGGCCACCAATATT 1212

Dd 1105 TTYCARGGNAAYWSNAAAYTYTMNGAYCYCCNGTNCARAAAYATTYTATHCNCNCTHGTN 1164

Qy 1213 GCACGTTTTATTAGATGAATCCTACCCAATGGCAGCAGAAAAATGTCATGCAAAATGGAG 1272

Dd 1165 GCNMGNATYTNMGNTNGTNCNCARACNTGCGCAYCARMGNATHGCTNYNAARGTNGAR 1234

Qy 1273 CTCGTCGGATGTCAGTTTTATTC 1294

Dd 1225 YTNATHGNTGYCARATHACNC 1246

RESULT 6

US-09-864-761-12562

; Sequence 12562, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 12562
;; LENGTH: 568
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC013497.4
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
US-09-864-761-12562

Query Match 10.2%; Score 232.4; DB 10; Length 568;
Best Local Similarity 99.6%; Pred. No. 9.4e-57;
Matches 233; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 140 AAGTGTGATGGTGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTTACATCCATAA 199
DB 231 AAGTGTGATGGTGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTTACATCCATAA 290
QY 200 ACTACCACAGACCTATCCCAACAGACACTGTTGTGTAATGGGAGATCCGTTGTAAGATGG 259
DB 291 ACTACCACAGACCTATCCCAACAGACACTGTTGTGTAATGGGAGATCCGTTGTAAGATGG 350
QY 260 GACAGAGAGTCCATCAAAATTTGGTGACTTTGACATTCGAGATTCGATTCCTGTCACCT 319
DB 351 GACAGAGAGTCCATCAAAATTTGGTGACTTTGACATTCGAGATTCGATTCCTGTCACCT 410
QY 320 TTAATTACTTGAGATTTATAATGGAATTTGGAGTGCAGCAGAACTGAAATAGGCA 373
DB 411 TTAATTACTTGAGATTTATAATGGAATTTGGAGTGCAGCAGAACTGAAATAGGTA 464

RESULT 7
US-09-864-761-29163
; Sequence 29163, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359

;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 29163
;; LENGTH: 228
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC013497.4
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
;; OTHER INFORMATION: NT HIT: D29810.1, EVALUE 1.00e-125
;; OTHER INFORMATION: EST_HUMAN HIT: A1565996.1, EVALUE 1.00e-125
;; OTHER INFORMATION: SWISSPROT HIT: P98063, EVALUE 1.00e-09
US-09-864-761-29163

Query Match 10.0%; Score 228; DB 10; Length 228;
Best Local Similarity 100.0%; Pred. No. 9.4e-56;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 143 GTGATGATGGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTTACATCCATAA 202
DB 1 GTGATGATGGGACACACTGTACTAGGCCCTGAGAGTGGAAACCCCTTACATCCATAA 60
QY 203 ACCCAGACCTATCCCAACAGACACTGTTGTGTAATGGGAGATCCGTTGTAAGATGGGAG 262
DB 61 ACCCAGACCTATCCCAACAGACACTGTTGTGTAATGGGAGATCCGTTGTAAGATGGGAG 120
QY 263 AGAGATTCGCATCAAAATTTGGTGACTTTGACATTTGAAGATTCGATTCCTGTCACITTA 322
DB 121 AGAGATTCGCATCAAAATTTGGTGACTTTGACATTTGAAGATTCGATTCCTGTCACITTA 180
QY 323 ATTACTTGAGAAATTTATAATGGAATTTGGAGTGCAGCAGAACTGAAATAG 370
DB 181 ATTACTTGAGAAATTTATAATGGAATTTGGAGTGCAGCAGAACTGAAATAG 228

RESULT 8
US-10-003-132-1
; Sequence 1, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.

[illegible]

Matches 191; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

```
OY 492 TCGTTATAGTAACAAAGATCAATTAATACCTGTTGGACACTGCATCCCAATTTTGGAAACC 551
Db 391 TCGAACAGATCAGTATGATTTAATAACCTGTTTGAAGCAGGACCCCAATTAATTCGAGGA 450
OY 552 TGAGTTTCAGTAACTAGTACGCCAGCTGGTGTCTGCTCTCTTTTGTGAGATATCTGGAAC 611
Db 451 AAATACAGCAATTCGCCAGCTGGCTGTAGAGACATAGCAGGAGATATTCTGGCAA 510
OY 612 AATTCCTCATGATATAGAGATTCCTGCCAATTCGCAATTCGATGGCTGTGCAATGAGGAGT 671
Db 511 TACAAAAGATGTTACAGAGATACCTTTTATTTGTGCAAGCTGCCATCCAGCAGGAT 570
OY 672 AGTGTCAACACAGCTTGGGGGCGCAATCAGTGTCTTAATTAGTAAGGATTTCCCTATTA 731
Db 571 CATCACAGATGAACATAGGTGGCCCATCAACTTGTCTCAGAGCAAGGGATAAGTCACTA 630
OY 732 TGAAAGTCTTTTGGCTAACACGCTCACATCTGTGTGGGACACTTATCTACAAGTCTTTT 791
Db 631 TGAAGGACTCTGGCAATGGCGTCTCTCCGGCATGTTCTTGTGCGAAAGCGATT 690
OY 792 TACATTTAAGACAAGTGGATTTATGGACACTGGGGATGGAGTCTGGTGTGATC 846
Db 691 TCTTTTACAAACCCAGGAATGAATATTACAACTGTGGCGATTCCATCAGTGATC 745
```

RESULT 14

```
US-09-974-298-167
; Sequence 167, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hwei-Wei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 167
; LENGTH: 4599
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 235636.1
US-09-974-298-167
```

Query Match 3.9%; Score 89.8; DB 9; Length 4599;
Best Local Similarity 50.7%; Pred. No. 4e-15;
Matches 242; Conservative 0; Mismatches 232; Indels 3; Gaps 1;

```
OY 810 ATGTTATGAACACTGGGATGAGTCTGGTGTGATCGCGGATCCTCAATAACAGCATC 869
Db 784 ATGCTCAGGCCCACTGGGAATTGAAGTGGAAATTATATCAAAACCAAGCAAAATCACAGCTTC 843
OY 870 ATCTGTCTGGAGTGTGACTGACCACACAGGGCAAGAGACACTTGGAAACCCCAAAAAGC 929
Db 844 CTCTACTC---ACCGAGCTCTTTTGGACTTCCAAAAAATGGTATCCCTACTATGCAGCTCT 900
OY 930 CAGGCTGAAAAAACCTTGGACCGCTTGGTGTCTTTGGCACTGATGAATACAGTGGTT 989
Db 901 TATAAGAGAGGGCTTTATAAATCGTGGACAGCTGCAGAAATGACAGATGGCCGTGGAT 960
OY 990 ACAAATAGATTTGAATGAAGAAAGAAAAATAACAGGCATTTAAACCACTGGATCCACCAT 1049
Db 961 TCAGATAAATTTGCAAAAGGAAATGAGAGTTACTGTGTGATTACCCAAAGGAGCCCAAGAG 1020
OY 1050 GGTGGAGCACAATTAATCTATCTGTCTGCCTACAGAAATCCCTGATGATGATGGGCAGAA 1109
Db 1021 GATTGGAAGCCAGAGTATATAAATTCCTACAAAATTTGCCTACAGTAATGATGGAAAGAC 1080
```

RESULT 15

```
US-09-880-107-2275
; Sequence 2275, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; TITLE OF INVENTION: Gene Logic, Inc.
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2275
; LENGTH: 6909
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M16967
US-09-880-107-2275
```

Query Match 3.8%; Score 86.6; DB 10; Length 6909;
Best Local Similarity 51.7%; Pred. No. 4.4e-14;
Matches 259; Conservative 0; Mismatches 224; Indels 18; Gaps 2;

```
OY 796 TTTAAGACAAGTGGATGTTATGGAACACTGGGATGAGTCTGTGTGATCGCGATCCT 855
Db 6271 TGTGAGTAAATGGATGTTCCACACCCCTGGGTATGGAATGGAAGATAGANAACAAG 6330
OY 856 CAAATAACAGCATCATCTGTGTGGAGTGGACTGACCACAGGCGCAAGAGAACAGTTGG 915
Db 6331 CAAATCACAGCTCTC-----GTTTAAGAAATCTTGGTGGGAGATTACTGG 6378
OY 916 AAACCCAAAAAGCCAGGCTGAAAAAACCTGGAC-----CGCCTGGGCTGCTTTGGC 969
Db 6379 GAACCTCTCCGTCGCGCTCTGAATGCCAGGACGCTGTAATCCCTGGCAAGCAAGGCA 6438
OY 970 ACTGATGAATACAGTGGTTTACAAATAGATTGTAATAAGAAAAAATAACAGGCATT 1029
Db 6439 AACACATAGCAGTGGCTAGAAATGATCTACTCAAGATCAAGAGATAAGCGCAATT 6498
OY 1030 ATAAACCACTGGATCCACCATGGTGGAGCAAAATTAATGTCTGTCTGCCCTACAGAAATCC 1089
Db 6499 ATACACAGGCTGCAAGTCTCTGTCTCTGAAATCTATGTAAGAGCTATACCATCCAC 6558
OY 1090 TACAGTATGATGGGCGAGAAATGGACTGTGTACAGAGACGCTGGTGTGGAGCAAGATAAG 1149
Db 6559 TACAGTGAAGGAGTGGAAATGGAAACCATACAGCTGAAATFCTCCATCGTGGGACAAG 6618
OY 1150 ATATTTCAAGGAACAAAGATTTATCACCAGGATGTCGGTAATAACTTTTGGCCCAACAAT 1209
Db 6619 ATTTTGAAGGAATAACTAATACCAAGGACATGTGAAGAACTTTTCAACCCCCCAATC 6678
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 21, 2003, 10:00:21 : Search time 75.1281 Seconds
(without alignments)
3875.409 Million cell updates/sec

Title: US-10-060-830-1114

Perfect score: 3470

Sequence: 1 MPLFLLLLLVLLLEDDAGA.....TQEVSGAGRGCDVFKKIL 653

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 393868 seqs, 222934149 residues

Total number of hits satisfying chosen parameters: 787736

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10060830/runat_16012003_091559_29042/app_query.fasta_1.1742
-DB=PublishedApplications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10060830.ecgn_1_1_86/grunat_16012003_091559_29042
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3422	98.6	5657	9	US-09-974-298-96 Sequence 96, Appl
2	897	25.9	640	10	US-09-879-536-480 Sequence 480, Appl
3	887.5	25.6	3151	9	US-10-003-132-1 Sequence 1, Appli
4	838	24.1	580	10	US-09-864-761-12400 Sequence 12400, A

5	828	23.9	467	10	US-09-864-761-30344 Sequence 30344, A
6	669.5	19.3	2145	9	US-10-003-132-12 Sequence 12, Appli
7	417	12.0	2836	9	US-10-003-132-3 Sequence 3, Appli
8	416	12.0	1871	10	US-09-823-038A-43 Sequence 43, Appl
9	355	10.2	4599	9	US-09-974-298-1675 Sequence 1675, Appl
10	353.5	10.2	6909	10	US-09-880-107-2275 Sequence 13, Appli
11	353.5	10.2	6925	9	US-10-115-563-13 Sequence 26, Appl
12	353.5	10.2	6925	9	US-10-115-563-26 Sequence 3020, Ap
13	346	10.0	1370	10	US-09-880-107-3020 Sequence 5, Appli
14	324	9.3	2868	9	US-10-003-132-5 GENERAL INFORMAT
15	322	9.3	4629	10	US-09-150-811-7 Sequence 1, Appli
16	322	9.3	9009	9	US-09-957-641-1 Sequence 3, Appli
17	320.5	9.2	7914	12	US-10-095-718-3 Sequence 14, Appl
18	317	9.1	970	10	US-09-215-450-14 Sequence 3, Appli
19	310.5	8.9	3404	12	US-10-104-440-3 Sequence 14, Appl
20	308.5	8.9	4999	9	US-10-007-968-14 Sequence 14, Appl
21	308.5	8.9	4999	10	US-09-740-211-14 Sequence 1, Appli
22	308.5	8.9	7944	12	US-10-095-718-1 Sequence 13, Appl
23	308.5	8.9	11933	9	US-10-007-968-13 Sequence 13, Appl
24	308.5	8.9	11933	10	US-09-740-211-13 Sequence 13, Appl
25	290	8.4	1509	9	US-10-003-132-13 Sequence 1, Appli
26	270.5	7.8	5653	12	US-10-104-440-1 Sequence 1, Appli
27	231	6.7	2490	10	US-09-900-518A-1 Sequence 61, Appl
28	220.5	6.4	3772	9	US-09-992-598-61 Sequence 61, Appl
29	220.5	6.4	3772	9	US-09-989-293A-61 Sequence 61, Appl
30	220.5	6.4	3772	9	US-09-989-735-61 Sequence 61, Appl
31	220.5	6.4	3772	9	US-09-990-444-61 Sequence 61, Appl
32	220.5	6.4	3772	9	US-09-989-730-61 Sequence 61, Appl
33	220.5	6.4	3772	9	US-09-990-436-61 Sequence 61, Appl
34	220.5	6.4	3772	9	US-09-991-181-61 Sequence 61, Appl
35	220.5	6.4	3772	9	US-09-993-687-61 Sequence 61, Appl
36	220.5	6.4	3772	9	US-09-989-734-61 Sequence 391, App
37	220.5	6.4	3772	9	US-10-028-072-391 Sequence 61, Appl
38	220.5	6.4	3772	9	US-09-997-653-61 Sequence 61, Appl
39	220.5	6.4	3772	10	US-09-989-722-61 Sequence 61, Appl
40	220.5	6.4	3772	10	US-09-989-723-61 Sequence 61, Appl
41	220.5	6.4	3772	10	US-09-989-279-61 Sequence 61, Appl
42	220.5	6.4	3772	10	US-09-989-727-61 Sequence 61, Appl
43	220.5	6.4	3772	10	US-09-989-731-61 Sequence 61, Appl
44	220.5	6.4	3772	10	US-09-989-732-61 Sequence 61, Appl
45	220.5	6.4	3772	10	US-09-991-073-61 Sequence 61, Appl

ALIGNMENTS

RESULT 1
US-09-974-298-96
; Sequence 96, Application US/0974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Ruel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 96
; LENGTH: 5657
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 241227.17
; NAME/KEY: unsure
; LOCATION: 4516
; OTHER INFORMATION: a, t, c, g, or other
US-09-974-298-96
Alignment Scores: 3.76e-310 Length: 5657
Pred. No.: 3.76e-310 Length: 5657


```

Db 2041 TCAGCCAGGCCAGTATGATACCCGAAAGCTGGGAAGCCAGGTCTACCTGCCCCAGAC 2100
QY 625 GluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaGlyArgAspGly 644
Db 2101 GAATGGTGTACCAGGTGGCCAGACACACAAGAAGTATCAGGAGCAGGAAGGATGGG 2160
QY 645 GluCysAspValPheLysGluIleLeu 653
Db 2161 GAATGGTGTATTTTAAGAAATCCYT 2187

RESULT 2
US-09-879-536-480/c
; Sequence 480, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/879,536
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/088,801
; PRIOR FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 480
; LENGTH: 640
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(640)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-480

Alignment Scores:
Pred. No.: 2,71e-75 Length: 640
Score: 897.00 Matches: 194
Percent Similarity: 91.98% Conservative: 1
Best Local Similarity: 91.51% Mismatches: 17
Query Match: 25.85% Indels: 6
DB: 10 Gaps: 0

US-10-060-830-1114 (1-653) x US-09-879-536-480 (1-640)
QY 299 LeuProPheIleAlaArgPheIleArgValAsnProThrGlnTrpGlnGlnLysIle 318
Db 630 TTGCCACCAATATTGGCAGC-TTTATTAGAT-AATCCTTACCCATGCCAGCAGAAATT- 574
QY 319 AlaMetLysMetGluLeuLeuGlyCysGlnPheIleProLysGlyArgProLysLeu 338
Db 573 GCCATGAAT- GAATGCTCGGATGTCAGTTATTCTAAGT-NGTCNTCCAAACTT 516
QY 339 ThrGlnProProProArgAsnSerAsnAspLeuLysAsnThrThrAlaProProLys 358
Db 515 ACTCAACCTCCACCTTTNGNAACAGCAATGNCCTCAAAACCACTACAGCCCTCCAAA 456
QY 359 IleAlaLysGlyArgAlaProLysPheThrGlnProLeuGlnProArgSerSerAsnGlu 378
Db 455 TA-GCCAAAGTGTGGCCCAAAATTTACGCAACCACTACAACTTCGCAGTAGCAATGAA 397
QY 379 PheProAlaGlnThrGluGlnThrThrAlaSerProAspIleArgAsnThrThrValThr 398

```

```

Db 396 TTTCTCTGCACAGACAGACAAACAACTGCCAGTCTGATATCAGAAATACCTACCGTAAC 337
QY 399 ProAsnValThrLysAspValAlaLeuAlaAlaValLeuValProValLeuValMetVal 418
Db 336 CCAAAATGTAACCAAGATGTAGCGCTGGCTGCGAGTTCTTCTCCCTGTGCTGGTCATGGTC 277
QY 419 LeuThrThrLeuIleLeuValCysAlaTrpPheIleTrpArgAsnArgLysLysLys 438
Db 276 CTCACCTACTCTCATTTAGTGTGCTTGGCACTGGAGAACACAGAAAGAAAA 217
QY 439 ThrGluGlyThrTyrAspLeuProTyrTrpAspArgAlaGlyTrpTrpLysGlyMetLys 458
Db 216 ACTGAAGGCACCTATGACTTACCTTACTGGGCGCGGCGAGTGTGGTGAAGGAATCAAG 157
QY 459 GlnPheLeuProAlaLysAlaValAspHisGluLeuThrProValArgTyrSerSer 478
Db 156 CAGTTTCTCTCTCAAAAGCAGTGGACCATGAGAAACCCAGTTCCTATACGACGC 97
QY 479 GluValAsnHisLeuSerProArgGluValThrThrValLeuGlnAlaAspSerAlaGlu 498
Db 96 GAAGTTAATCAGCTGAGTCCAGAGAAGTCCACACAGTGTGCAGGCTGACTCTGCAGAG 37
QY 499 TyrAlaGlnProLeuValGlyGlyIleValGlyThr 510
Db 36 TATGCTCAGCCACTGTGTAGGAGGAATTGTGTGGTACC 1

RESULT 3
US-10-003-132-1
; Sequence 1, Application US/10003132
; Publication No. US20020192750A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Gao, Zeren
; APPLICANT: Shoemaker, Kimberly E.
; TITLE OF INVENTION: NEUROFILIN HOMOLOG 2CUBS
; FILE REFERENCE: 00-62
; CURRENT APPLICATION NUMBER: US/10/003,132
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/249,004
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (76)...(2223)
US-10-003-132-1

Alignment Scores:
Pred. No.: 2.15e-73 Length: 3151
Score: 887.50 Matches: 230
Percent Similarity: 50.76% Conservative: 103
Best Local Similarity: 35.06% Mismatches: 221
Query Match: 25.58% Indels: 102
DB: 9 Gaps: 21

US-10-060-830-1114 (1-653) x US-10-003-132-1 (1-3151)
QY 13 LeuLeuLeuGluAspAlaGlyAlaGlnGlnGlyLysTyrCysGlyLeuGlyLeuGlnMet 32
Db 370 CTTCTTCACACAGCTCTTCAGATCAATATGCTCCACTACTGTGGA--AGTATGACTGTT 426
QY 33 AsnHisSerIleGluSerLysGlyAsnGluIleThrLeuLeuPheMetSerGlyIleHis 52
Db 427 CCCAAGACACTCTTCTTGAACACACAGTCAAGTAACCGTCCGCTTTTGAGAGTGGATCCAC 486
QY 53 ValSerGlyArgGlyPheLeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThr 72
Db 487 ATTTCTGCCGGGGTTTTTTTGTGCTACCTATGCGACGACGACCATCCAGATTATATAACA 546

```

QY 73 CysLeuAspThrAlaSerAsnProGluPheSerLysTyrCysProAlaGly 92
Db 547 TGTGTGAACGAGCTACCATATTGGAACAGACAGATACAAATCTCCCGAGCTGGT 606
QY 93 CysLeuLeuProPheAlaGluSerGlyThrIleProHisGlyTyrArgAspSer 112
Db 607 TGTAGAGACCTAGCAGGAGACATTTCTGGGAATATGTAGATGATATAGATACCTCT 666
QY 113 ProLeuCysMetAlaGlyValHisAlaGlyValValSerAsnThrLeuGlyGlnIle 132
Db 667 TTATTGTGCAAGCTGCATCCATGAGGAATATGCTGATGAATAGTGCCAGATC 726
QY 133 SerValValIleSerLysGlyIleProTyrGluSerSerLeuAlaAsnValThr 152
Db 727 AGTGTGCTTCAGCCCAAGGATCAGTCGATATGAAGGATTTCTGGCCATGTTCTT 786
QY 153 SerValValGlyHisLeuSerThrSerLeuPheThrPheLysThrSerGlyCysTyrGly 172
Db 787 TCGAGGATGGTTCCTCTGTGACACAAGCATTTCTGTTTACCTCCATGTTGTCAGCAGA 846
QY 173 ThrLeuGlyMetGluSerGlyValIleAlaAspProGlnIleThrAlaSerValLeu 192
Db 847 TCTTGTGATTTTGAA-----CCTGACGGGCAATCAGAGCTTCTTCTCATGG 894
QY 193 GluTrpThrAspHisThrGlyGlnIleAsnSerTrpLysProLysLysAlaArgLeuLys 212
Db 895 CAGTCGGTCAATGATGAGGAGACCAAGTTCAGTGGTCTCTGGCCCAAGCCGACATTCAG 954
QY 213 LysProGlyProProTyrAlaAlaPheAlaThrAspGlu-----TyrGlnTrp 228
Db 955 GACCAAGGCCCATCATGGCTTCGGGCGACAGTAGCAACAACCAACACAGAGAGTGG 1014
QY 229 LeuGlnIleAspLeuLysGlyLysIleThrGlyLysIleThrGlySerThr 248
Db 1015 CTGAGATCGATTGGGGGAGAAAGAAATAACAGGAATAGGACACAGGATCTTACA 1074
QY 249 MetValGluHisAsnTyrValSerAlaTyrArgIleLeuTyrSerAspGlyGln 268
Db 1075 CAGTCGAACCTCACTTTTAAAGATTTGTGTGATGACTTCAAAAACAATAATCT 1134
QY 269 LysTrpThrValTyrArgGluProGlyValGlnAspLysIlePheGlnIleAsnLys 288
Db 1135 AAGTGAAGACCTATAAAGGAATTTGTAATAATGAAGAAAGGTGTTTTCAGGTAACCT 1194
QY 289 AspTyrHisGlnAspValArgAsnAspPheLeuProLleIleAlaArgPheIleArg 308
Db 1195 AACTTTCGGGACCCAGTGCATAAACAATTTTCATCCTCCCATGCTGGCCAGATATGTCGG 1254
QY 309 ValAsnProThrGlnTrpGlnGlnLysIleAlaMetLysMetGluLeuLeuGlyCysGln 328
Db 1255 GTTGTCCCGGACAGATGCCACGAGGATACCTTTGAAGGTGGAGCTCATTTGGTTCGAC 1314
QY 329 PheIleProLysGlyArgProLysLeuThrGlnProProProProArgAsnSer 348
Db 1315 -----ATTACACA-----GGTAAT 1329
QY 349 Asp-----LeuLysAsnThrThrAlaProProLysIleAlaLysGlyArgAla 364
Db 1330 GATTCAATTGGTGGCGCAAGCAAGTCAAGACCACTGTTTCAACTAAGAAAGAGAT 1389
QY 365 ProLysPheThrGlnProGlnProArgSerSerAsnGluPheProAlaGlnThrGlu 384
Db 1390 GAGCAATCAAGGCCCATC-----CCCTCGGAA 1419
QY 385 GlnThrThrAlaSerProAspIleArgAsnThrThrValThrProAsnValThrLysAsp 404
Db 1420 GAAACA-----TCCACGAGGAATAACATTACACG----- 1449
QY 405 ValAlaLeuAlaValLeuValProValLeuValMetValLeuThrLeuIleLeu 424
Db 1450 GTGGCTATTCCATTGGTCTCTCTGTTGCTGGAATGGGATCTTTTGCA 1509

QY 425 IleLeuValCysAlaTrpHisTrpArgAsnArgLysLysThrGluGlyThrTyrAsp 444
Db 1510 GCCTTT-----AGAAAGAAGAAGAAGGAAGT----- 1539
QY 445 LeuProTyrTrp-----AspArgAlaGlyTrpTrpLysGlyMetLysGlnPhe 460
Db 1540 ---CCGTATGATCAGCAGAGGCTCAGAAAACAGACTGTTTGAAGCAGATTAATATATCC 1596
QY 461 LeuProAlaLysAlaValAspHisGluLeuThrProValArgTyrSerSerGlu--- 479
Db 1597 TTT-----GCCACATCAGCTCAGCTGAGTTT 1623
QY 480 ---ValAsnHisLeuSerProArgGluValThr-----ThrValLeuGlnAlaAsp 495
Db 1624 ACCATCAGCTATGATATGAGAGGAGATGACACAAAAGTTAGATCTCATCACAAGTAT 1683
QY 496 SerAlaGluTyrAlaGlnProLeuValGlyIleValGlyThrLeuHisGlnArgSer 515
Db 1684 ATGCAGATTTACCAGCAGCCCTCATGTTGGCAGGACAGTCAAGGAGAGGCTCC 1743
QY 516 ThrPheLysPro-----GluGluGlyLysGluAlaGlyTyrAlaAspLeuAspPro--- 532
Db 1744 ACCTTCGGGCCCATGAGACAGGATGCCGAGGAGCAGGG---GTGAGCACCGATGCCGCG 1800
QY 533 -----TyrAsnSerProGlyGlnGluValTyrHisAlaTyrAlaGluProLeuProIle 550
Db 1801 GGCACATATGACTCCCGCGACGGCGCGCCGACGAGTACGCGCTCCCTGGCGGCC 1860
QY 551 ThrGlyProGluTyrAlaThrProIleMetAspMetSerGlyHisProThrThrSer 570
Db 1861 CCGAGCGCCGAGTACGCGACGCCCATCGTG-----GAGCGCAGCTGTGCGCGCC 1911
QY 571 ValGlyGlnProSerThrSerThrPheLysAlaThrGlyAsnGlnPro----- 586
Db 1912 CACACGTTCTCTCGCGAGAGCGCTACCGCTCCCGAGGCCCGCCGCGCCACAAACAC 1971
QY 587 -----ProLeu-----ValGlyThrTyrAsnThrLeuLeu 597
Db 1972 TCCTCTCTCGCGCGCTTCTCCCGTAGCGGTGTGGGCGCCGAGGAGACATAT 2031
QY 598 SerArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAla-Glyly 617
Db 2032 CAAAGGCCACACGCGACAGCCCTCGGACAGGGCTACGACCGCGCCCAAGCTGTACG 2091
QY 617 sProGlyLeuProAlaProAspGluLeuValTyrGlnValProGln 632
Db 2092 GCCTTCGCCACCGAAAGCGGACACCTGACTCTCAGAAAGCCCA 2137

RESULT 4

US-09-864-761-12400
; Sequence 12400, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL I
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 12400
;; LENGTH: 580
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC013497.4
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.47
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
US-10-060-830-1114 (1-653) x US-09-864-761-12400 (1-580)
Alignment Scores:
Pred. No.: 7,77e-70 Length: 580
Score: 838.00 Matches: 161
Percent Similarity: 95.38% Conservative: 4
Best Local Similarity: 93.06% Mismatches: 7
Query Match: 24.15% Indels: 1
DB: 10 Gaps: 0
US-10-060-830-1114 (1-653) x US-09-864-761-12400 (1-580)
QY 482 HisLeuSerProArgGluValThrValLeuGlnAlaasp-SerAlaGluTyrAlaG1 501
Db 21 AATCTCTTACTAGTAAAGACTTTCTGTTGTTGTTGATATCCACAGAGTATGCTCA 80
QY 501 nProLeuValGlyGlyLeuValGlyThrLeuHisGlnArgSerThrPheLysProGluG1 521
Db 81 GCCACTGGTAGGAGGAATTCTTGTCACACTTCATCAAGATCTACCTTTAAACAGAGA 140
QY 521 uGlyLysGluAlaGlyTyrAlaaspLeuaspProTyrAsnSerProGlyGlnGluValty 541
Db 141 AGGAAAGRAGCAGGCTATGCAGACCTAGATCTCTTACAACTCCACAGGCGAGGAGTTTA 200
QY 541 rHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrProIleLeMe 561
Db 201 TCATGCTATGCTGAACCACTCCCAATTACGGGCGCTGAGTATCAACCCCAATCATCAT 260
QY 561 tAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThrPheLysAl 581
Db 261 GGACATGTCAGGCGACCCACACTTCAGTTGGTCAGCCCTCCACATCCACTTCAAGGC 320
QY 581 aThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSerArgThrAs 601
Db 321 TACGGGGAACCAACCTCCCCCACTAGTGGGAACCTTACAATACACTTCTCTCCAGGACTGA 380

QY 601 pSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysProGlyLeuPr 621
Db 381 CAGCTGCTCTCAGCCAGCCAGTATGATACCCCGAAGCTGGGAGCCAGGTCTACC 440
QY 621 oAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSerGlyAlaG1 641
Db 441 TGGCCAGACGCAATTTGGTGTACCCAGGTGCCACAGACACACAGAGTATCATCAGGAGCAGG 500
QY 641 YArgAspGlyGluCysAspValPheLysGluIleLeu 653
Db 501 AAGGATGGGGAATGTGATGTTTTTAAGAAATCCCTT 537
RESULT 5
US-09-864-761-30344
;; Sequence 30344, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
;; FILE REFERENCE: Aeomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263,6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 30344
;; LENGTH: 467
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC013497.4
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.61
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.47

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.46
OTHER INFORMATION: EST_HUMAN HIT: AW583777.1, EVALUUE 3.00e-51
OTHER INFORMATION: NT HIT: Z70177.1, EVALUUE 2.90e-01
OTHER INFORMATION: SWISSPROT HIT: Q9Y011, EVALUUE 7.40e-01
US-09-864-761-30344

Alignment Scores:
Pred. No.: 4,88e-69 Length: 467
Score: 828.00 Matches: 155
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23,86% Indels: 0
DB: 10 Gaps: 0

US-10-060-830-1114 (1-653) x US-09-864-761-30344 (1-467)

Qy 499 TyrAlaGlnProLeuValGlyLeuValGlyThrLeuHisGlnArgSerThrPheLys 518
Db 3 TATGCTCAGCCACGCTGAGGAGGAAATGTTGGTACACTTATCAACAGATCTACCTTTAAA 62
Qy 519 ProGluGluGlyLysGluAlaGlyTyrAlaAspLeuAspProTyrAsnSerProGlyGln 538
Db 63 CCAGAAGAAGAAAGAGGAGGCTATGCGACACCTAGATCCTTACAACCTCACCAGGCGAG 122
Qy 539 GluValTyrHisAlaTyrAlaGluProLeuProIleThrGlyProGluTyrAlaThrPro 558
Db 123 GAAGTTTATCATGCTATGCTGCAACCACTCCCAATTCAGGGGCGCTGAGTATGCAACCCCA 182
Qy 559 IleIleMetAspMetSerGlyHisProThrThrSerValGlyGlnProSerThrSerThr 578
Db 183 ATCATCATGACATGTCAGGCGACCCACACACTTCAGTTGCTCAGCCCTCCACATCCACT 242
Qy 579 PheLysAlaThrGlyAsnGlnProProLeuValGlyThrTyrAsnThrLeuLeuSer 598
Db 243 TTCAGGCTACGGGGAACCAACCTCCCCCACTAGTGGGAACCTTACAATACACTTCTCTCC 302
Qy 599 ArgThrAspSerCysSerSerAlaGlnAlaGlnTyrAspThrProLysAlaGlyLysPro 618
Db 303 AGGACTGACAGCTGCTCTCAGCCAGGCGCCAGTATGATACCCCGAAGCTGGGAAGCCA 362
Qy 619 GlyLeuProAlaProAspGluLeuValTyrGlnValProGlnSerThrGlnGluValSer 638
Db 363 GGTCTACCTGCCCCAGACGATGTTGTTACCGAGTGCACAGGCGCACACAGAGATATCA 422
Qy 639 GlyAlaGlyArgAspGlyGluCysAspValPheLysGluIleLeu 653
Db 423 GGAGCAGGAAGGATGGGGAATGTGATGTTTAAAGAAATCCTT 467

RESULT 6

US-10-003-132-12
Sequence 12, Application US/10003132
Publication No. US20020192750A1

GENERAL INFORMATION:

APPLICANT: Fox, Brian A.
APPLICANT: Shoemaker, Kimberly E.
TITLE OF INVENTION: NEUROFILIN HOMOLOG ZCUB5
FILE REFERENCE: 00-62
CURRENT APPLICATION NUMBER: US/10/003,132
PRIOR FILING DATE: 2001-11-15
PRIORITY DATE: 2000-11-15
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 12
LENGTH: 2145

ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: degenerate nucleotide sequence
NAME/KEY: misc.feature
LOCATION: (1)...(2145)
OTHER INFORMATION: n = A,T,C or G

US-10-003-132-12

Alignment Scores:

Pred. No.: 3,05e-53 Length: 2145
Score: 669.50 Matches: 181
Percent Similarity: 40.22% Conservative: 76
Best Local Similarity: 28.33% Mismatches: 287
Query Match: 19.29% Indels: 95
DB: 9 Gaps: 19

US-10-060-830-1114 (1-653) x US-10-003-132-12 (1-2145)

Qy 21 GlnGlnGlyLysTyrCysGlyLeuGlnMetAsnHisSerIleGluSerLysGly 40
Db 319 CARTAYGNCNCNTATGYGGN-----WSNATGACNGTNCNNAARGARTNTNTYN 369
Qy 41 Asn-----GluIleThrLeuLeuPheMetSerGlyIleHisValSerGlyArgGlyPhe 58
Db 370 AAYACNWSNGARGTACNGTNGNTTYGARWSNGCNSNCAVATHTWSNGNGMGNGNTTY 429
Qy 59 LeuAlaSerTyrSerValIleAspLysGlnAspLeuIleThrCysLeuAspThrAlaSer 78
Db 430 YNTNTNACNTAYGCNWSNWSNGAYCAYCCNGAYTNTNATHACNTGYTTNGARMGNCNWSN 489
Qy 79 AsnPheLeuGluProGluPheSerLysTyrCysProAlaGlyCysLeuLeuProPheAla 98
Db 490 CAYTAYTNAARACNGARTATWSNNAARTTYTTCNGCNGGNTGYMGNGAYGTNGCNGN 549
Qy 99 GluIleSerGlyThrIleProHisGlyTyrArgAspSerSerProLeuCysMetAlaGly 118
Db 550 GAYATHWSNGGNAAYATGTTNGAYGNTAYMGAYACNWSNTVNTNTGYAARGCNGCN 609
Qy 119 ValHisAlaGlyValValSerAsnThrLeuGlyGlyGlnIleSerValIleSerLys 138
Db 610 ATHCAYCGNGGNATHGCGNGAYGARYTNGGNGCNCARATHWSNGTNTNCARMGNAAR 669
Qy 139 GlyIleProTyrTyrGluSerSerLeuAlaAsnValThrSerValValGlyHisLeu 158
Db 670 GGNATHWSNMGTAYTGARGGNATHYTGNCNAAAYGGNTNTWSNMNGNGAYGNGNSNYTN 729
Qy 159 SerThrSerLeuPheThrPheLysThrSerGlyCysTyrGlyThrLeuGlyMetGluSer 178
Db 730 WSNAGYAAARMGNTTYTNTTYACNWSNAAAYGGNTGYWSNMGNWSNTNNTTYGAR--- 786
Qy 179 GlyValIleAlaAspProGlnIleThrAlaSerSerValLeuGluThrPheThrAspHisThr 198
Db 787 -----CCNGAYCGNCARATHMGNGCNSNWSNWSNTGGCARSNGTNAAYGARWSN 837
Qy 199 GlyGlnGluAsnSerTriLysProLysAlaArgLeuLysLysProGlyProTriP 218
Db 838 GGNAGYARGTNCAYTGGWSNCCNGCNGCARGCMGNTNCARGAYCARGCNGCNSNTGG 897
Qy 219 AlaAlaPheAlaThrAspGlu-----TyrGlnTriLeuGlnIleAspLeuAsn 234
Db 898 GCNWSNGNGAYWSNWSNAAAYAAAYAACCCCMGNGARTGGTYNGARATHGAYTYNGGN 957
Qy 235 LysGluLysLysIleThrGlyIleIleThrThrGlySerThrMetValGluHisAsnTyr 254
Db 958 GARAARAARAARATHACNGGNATHMGNCACNGGWSNACNACNARSNAAYTYTYAATY 1017
Qy 255 TyrValSerAlaTyrArgIleLeuTyrSerAspGlyGlnLysTriPThrValTyrArg 274
Db 1018 TAYGTNARWSNTTYCTNATGAATYTAARAAYAAAYAAAYWSNAARTGGARACNTAYAR 1077
Qy 275 GluProGlyValGluGlnAspLysIlePheGlnGlyAsnLysAspTyrHisGlnAspVal 294
Db 1078 GGNATHGTNAAAYAAAYGARGAARGTNTTYCARGGNAAYWSNAAYTYTYMGNGAYCCNGTN 1137
Qy 295 ArgAsnAsnPheLeuProIleIleAlaArgPheIleArgValAsnProThrGlnTriP 314
Db 1138 CARAAAYATTYATHTCCNCCNATHGTNGCNGMNTATYGTNMNGNTGTNCCNACARACNTGG 1197
Qy 315 GlnGlnLysIleAlaMetLysMetGluLeuGlyCysGlnPheIleProLysGlyArg 334